

Study of Beauty Photoproduction with the ZEUS Experiment at the Electron-Proton Collider HERA

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Frau Dipl.-Phys. Ana Gloria Yagües Molina
geboren am 07.09.1979 in Madrid

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Dr. h.c. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:

Prof. Dr. Christian Limberg

Gutachter:

1. Prof. Dr. Thomas Lohse
2. Priv.-Doz. Dr. Achim Geiser
3. Prof. Dr. Heiko Lacker

eingereicht am: 28 November 2007

Tag der mündlichen Prüfung: 22. Februar 2008

Abstract

This thesis presents a measurement of beauty photoproduction in ep interactions at the HERA collider working at a center of mass energy of $\sqrt{s} = 318$ GeV. During the HERA luminosity upgrade period 2000/2001 the tracking system of the ZEUS detector was enhanced with a silicon Micro Vertex Detector (MVD). The implementation of the MVD provides high precision measurements that allow new identification techniques of the beauty quarks based on their heavy mass and long lifetime. Two goals are followed in this thesis: first the development of a reliable algorithm (secondary vertex b -tagging) to identify jets originating from b quarks by exploiting exclusively the full MVD potential and the properties of the b quark, and second the determination of its performance by obtaining first measurements of inclusive beauty dijet photoproduction.

The main result presented here is based on a data set collected in 2004 corresponding to an integrated luminosity of 35 pb^{-1} . Events containing at least two jets, each having an identified secondary vertex, are selected. The fraction of events containing b -quarks is extracted using the characteristic distributions of reconstructed decay length and invariant mass of the secondary vertex. The jets are required to be within the pseudorapidity region $|\eta| < 1.5$ with transverse momenta $p_T > 7(6) \text{ GeV}/c$ respectively. No requirements are imposed on the b decay final state so that the measurement is kept inclusive. This is the first such measurement in ZEUS.

Total and differential cross sections as a function of transverse momentum $d\sigma/dp_T$ and pseudorapidity $d\sigma/d\eta$ of the b -jet are measured. The obtained results are compared with a previous published measurement carried out with a HERA-I data set (luminosity ≈ 3 times larger) in which the beauty signal is extracted by a completely independent method (P_T^{rel} method in semileptonic μ -decays). Furthermore, the results on $d\sigma/dp_T$ were extrapolated to obtain a measurement at the b -quark level. This measurement is compared with all previous ZEUS and H1 results. Good agreement is found. The calculated differential cross sections are finally compared with leading order (LO) plus parton-shower Monte Carlo simulations as well as next-to-leading order (NLO) QCD theoretical predictions and found to be compatible within the statistical and systematic errors.

Motivated by the results of this exploratory analysis of beauty quark production using secondary vertex b -tagging and profiting from improvements of

the ZEUS tracking software since this first analysis, a more precise measurement became possible. The new analysis, still ongoing, is based on the data set collected in 2005 by the ZEUS detector and profits of higher statistics ($L = 132.6 \text{ pb}^{-1}$) as well as of an improved detector alignment and calibration and a better understanding of the tracking/vertexing procedures which contribute significantly to the reduction of the uncertainties of the measurements. Though there is still room for further improvements, the higher precision achieved so far can be already seen in the first preliminary results presented in the second part of this work. Moreover, thanks to the higher statistics and improved performance, the measurement of $d\sigma/dp_T$ could be extended to larger p_T values, not reached before by other analyses.

Keywords:

beauty-production, secondary vertex tagging, decaylength, MVD ZEUS

Zusammenfassung

In dieser Dissertation wird eine Messung von Beauty-Produktion im Bereich der Photoproduktion in ep Kollisionen am HERA-Speicherring bei $\sqrt{s} = 318$ GeV beschrieben. Während des HERA-Umbaus zur Steigerung der Luminosität 2000/2001 wurde der neue Silizium-Mikrovertex-Detektor (MVD) in das ZEUS-Experiment eingebaut. Die Auflösung des MVD ermöglicht Präzisionsmessungen mit denen sich neue Methoden zur Identifikation der Beauty-Quark entwickeln lassen. Diese Methoden nutzen sowohl die große Masse der b -Quarks als auch deren lange Lebensdauer aus. Zwei Ziele werden in dieser Arbeit verfolgt: erstens die Entwicklung eines zuverlässigen Verfahrens (Sekundärvertex-Tagging-Methode) zur Identifikation von Jets die aus b -Quarks stammen unter Ausnutzung der MVD-Information und Eigenschaften der Beauty-Quarks, und zweitens die Bestätigung der Anwendbarkeit dieser Methode durch erste Messungen der inklusiven Beauty-Photoproduktion. Beide Ziele wurden erreicht.

Die Messung verwendet einen vom ZEUS-Detektor aufgenommenen Datensatz mit einer integrierten Luminosität von 35 pb^{-1} . Selektiert werden Ereignisse mit mindestens zwei Jets, jeder davon mit einem rekonstruierten Zerfallsvertex (auch Sekundärvertex genannt). Die Jets müssen im Pseudorapiditätsbereich $|\eta| < 1.5$ mit Transversalimpulsen $P_T > 7(6) \text{ GeV}/c$ liegen. Der Anteil der Ereignisse die Beauty-Quarks enthalten wurde mit Hilfe der charakteristischen Verteilung der Zerfallslänge und der invarianten Masse der Zerfallsvertex bestimmt. Dies ist die erste solche Analyse in ZEUS.

Totale und differentielle Wirkungsquerschnitte als Funktion des Transversalimpulses $d\sigma/dp_T$ und der Pseudorapidität $d\sigma/d\eta$ der Jets wurden gemessen. Alle Ergebnisse werden mit der vorher publizierten Messung verglichen. Diese Messung basierte auf drei mal höherer Luminosität und benutzt eine komplett unabhängige Methode zur Bestimmung des Anteils von Beauty-Ereignissen (P_T^{rel} Methode in semileptonischen μ -Zerfällen). Weiterhin wurden die $d\sigma/dp_T$ Ergebnisse extrapoliert um eine Messung auf b -Quark-Niveau zu erhalten. Der Vergleich wird mit allen vorhandenen Ergebnisse von ZEUS und H1 ergibt gute Übereinstimmung. Unter Berücksichtigung der statistischen und systematischen Unsicherheiten sind die gemessenen Wirkungsquerschnitte sowohl konsistent mit Monte-Carlo-Simulationen in führender Ordnung (LO) plus Partonschauer als auch mit QCD Vorhersagen in nächstführender Ordnung (NLO).

Aufgrund der vielversprechenden Ergebnisse dieser ersten Analyse und der mittlerweile erzielten Verbesserungen der Ereignis-Rekonstruktion, wurde eine zweite präzisere Messung der Beauty-Produktion unter Benutzung der Zerfallsvertex-Methode als Identifikation der Beauty-Ereignisse möglich. Die noch laufende Analyse profitiert von größerer Statistik (Datensatz von 2005 mit integrierter Luminosität von 132.6 pb^{-1}) sowie einer verbesserten Kalibrierung des Detektors. Die während der ersten Analyse hinzugewonnenen Kenntnisse über Spur- und Vertexrekonstruktionsverfahren werden ebenfalls zu einer Reduzierung der systematischen Unsicherheiten beitragen. Es sind zwar noch weitere Verbesserungen möglich, jedoch ist die bisher erreichte höhere Präzision bereits aus den im zweiten Teil dieser Arbeit vorgestellten vorläufigen Ergebnissen ersichtlich. Zudem konnten die Messungen dank erhöhter Statistik und höherer Effizienz auch auf größere p_T -Werte ausgedehnt werden.

Schlagwörter:

Beauty-Produktion, Sekundärvertex-Tagging, Zerfallslänge, MVD ZEUS

To the memory of Nikolaj Pavel

List of Figures

2.1	Left: Electron-proton scattering in Neutral Current (NC). Right: Electron-proton scattering in Charged Current (CC) . . .	7
2.2	Leading order direct photon processes: (Left) Boson-Gluon fusion and (right) QCD Compton	14
2.3	Leading order resolved photo processes: (Left) Boson-Gluon fusion and (right) QCD Compton	16
2.4	Space-time development of the $q\bar{q}$ state in breakup situation .	19
2.5	Diagram for B^- decays in the spectator model.	22
2.6	Schematic structure of ep generator in a BGF event.	27
2.7	Differential branching ratio dB/dp for the electron momentum by Belle and BaBar	28
2.8	Real NLO QCD contributions to heavy quark production . . .	28
2.9	Virtual NLO QCD contributions to heavy quark production .	29
3.1	Observation of dimuon resonance at 9.5 GeV: the Υ meson. . .	31
3.2	Beauty production cross sections in pion (a) and proton (b) collisions.	32
3.3	Beauty quark cross section from $D\emptyset$ and CDF experiments. . .	34
3.4	CDF beauty quark preliminary results for $\sqrt{s} = 1.96 \text{ TeV}$. .	35
3.5	Differential beauty cross sections as function of the pseudora- pidity and the transverse momentum of the b -jet.	36
3.6	Schematic illustration of the impact parameter in the $r\phi$ -plane.	37
3.7	Beauty tagging distributions from H1 beauty photoproduction measurement in dijets events.	38
3.8	Beauty photoproduction cross section measurements from H1.	40
4.1	Airplane view of the DESY site in Hamburg.	41
4.2	Layout of HERA and pre-accelerators.	43
4.3	Integrated luminosity delivered by HERA for the running pe- riods 1993-200 and 2002-2005	45
4.4	The ZEUS coordinate system	46
4.5	Cross section of the ZEUS detector along the beam axis	47

4.6	Cross section of the ZEUS detector perpendicular to the beam axis	48
4.7	A x-y cross section of one octant of the CTD.	49
4.8	$x - y$ view through CTD and layout of a drift cell in the CTD	50
4.9	Longitudinal MVD section	52
4.10	BMVD section	53
4.11	MVD: half modules and ladders	54
4.12	FMVD wheels	54
4.13	Schematic view of the UCAL along the beam axis.	57
4.14	FCAL module and tower structure in FCAL and RCAL	58
4.15	Schematic view of the ZEUS luminosity monitor system.	60
4.16	The ZEUS trigger and the data acquisition system.	61
5.1	General structure of the LO plus parton shower MC event generator.	64
5.2	Schematic diagram of the ZEUS analysis and MC reconstruction chain.	67
6.1	Helix parametrization in the XY plane of a positively charged track	73
6.2	Example of a Kalman prediction step.	75
6.3	Resolution of single particles MC simulations	78
6.4	Schematic plot of the island clustering algorithm	79
6.5	Map of dead material distribution	81
7.1	Relation between \mathbf{q}_i and (V, p_i) in a vertex fit.	86
7.2	Perigee parameters in the x-y projection.	89
7.3	HERA and H1 vertical and horizontal beam position measurements for 2004 data.	93
7.4	Dependence on Z of the primary vertex coordinates X and Y for 2004 data.	94
7.5	Resolution of the secondary vertex coordinates in a $b\bar{b}$ direct PHP MC sample for 2004 data.	95
8.1	Schematic representation of a secondary vertex inside a jet	101
8.2	Distance ΔR between tracks and jet in the (η, ϕ) plane	103
8.3	Schematic representation of the signed transverse decay length L_{xy}	105
8.4	Decay length distributions for beauty, charm and light flavour samples.	106
8.5	Superimposed decay length distributions of beauty, charm and light flavour samples.	107

8.6	Control distributions of the track related variables	109
8.7	Control distributions of the jet related variables	110
8.8	Control distributions of the MVD-CTD variables	111
9.1	2D plot of the χ^2 distribution for $0.5 < p_T < 1 \text{ GeV}$	114
9.2	Distribution of the Breit-Wigner gamma (top) and Gaussian sigma (bottom) in bins of p_T	116
9.3	Comparison of the decay length distribution for MC and data before and after smearing.	117
9.4	Secondary vertex mass M_{SV} distributions.	118
9.5	$(L_{xy}^+ - L_{xy}^-)$ distributions in bins of M_{SV}	120
9.6	Global $(L_{xy}^+ - L_{xy}^-)$ fit.	122
9.7	$(L_{xy}^+ - L_{xy}^-)$ distributions in bins of p_T after fit.	123
9.8	$(L_{xy}^+ - L_{xy}^-)$ distributions in bins of η after fit.	124
10.1	Differential cross sections as a function of the transverse mo- mentum and the pseudorapidity of the jet.	130
11.1	Differential cross sections for b -quark as a function of the b -quarks transverse momentum.	133
12.1	Display of trimuon beauty candidate with ZeVis.	137
12.2	2005 Control distributions of the track related variables	140
12.3	2005 Control distributions of the jet related variables	141
12.4	2005 Control distributions of the MVD-CTD variables	142
12.5	$(L_{xy}^+ - L_{xy}^-)$ - distribution before and after χ^2 cut.	143
12.6	The MC transverse decay length, error and significance distri- butions compared to the data for 2005 data.	145
12.7	The MC transverse decay length and decay length significance distributions compared to the data for 2005 data after MVD hits smearing.	147
12.8	Radar plot of MVD alignment (azimuthal angle zones) and azimuthal angle distribution divided by zones.	147
12.9	Transverse decay length in bins of azimuthal angle.	148
12.10	Decay length significance in bins of azimuthal angle.	149
12.11	Transverse decay length L_{xy} and significance S distributions in bins of the transverse momentum.	150
12.12	Subtracted distribution $(S^+ - S^-)$	151
12.13	Transverse decay length with respect to the reduced primary vertex.	153
12.14	M_{SV} and $(S^+ - S^-)$ -distributions in three bins of M_{SV}	154
12.15	Global $(S^+ - S^-)$ fit	159

12.16	Global ($S^+ - S^-$) fit before redefinition of L_{xy}	160
12.17	($S^+ - S^-$) distributions in bins of p_T after fit.	161
12.18	($S^+ - S^-$) distributions in bins of η after fit.	162
12.19	2005 differential cross sections as a function of the transverse momentum and the pseudorapidity of the jet.	163
12.20	Comparisons of the differential cross sections as a function of the transverse momentum and the pseudorapidity of the jet for 2004 and 2005 data	164
12.21	2005 Differential cross sections as a function of the transverse momentum of the jet (wide p_T range).	164
12.22	Differential cross sections for b -quark as a function of the b -quark transverse momentum.	166
A.1	Schematic cross section of a silicon detector.	168
A.2	Energy band structure of conductors, insulators and semicon- ductors.	169
B.1	Dependence of the shift on Z and θ	174
B.2	Transverse decay length distribution after 2D vertexing.	175

List of Tables

2.1	The fundamental particles in the Standard Model and their interactions.	6
2.2	Examples of B -hadrons	22
3.1	Collection of beauty production measurements at fixed target experiments and collider experiments $p\bar{p}$	32
3.2	Tevatron running periods.	33
4.1	HERA design parameters	44
5.1	MC samples used for the 2004 and 2005 data analysis I.	69
7.1	B and E matrices	89
8.1	MC samples used for the 2004 data analysis	98
8.2	Selection criteria for tracks used in the b-tagging algorithm. . .	102
8.3	Selection criteria for jets used in the b-tagging algorithm. . . .	102
10.1	Differential cross section for b-jets as a function of p_T^{b-jet}	127
10.2	Differential cross section for b-jets as a function of η^{b-jet}	127
10.3	Summary of the systematic errors in 2004 data analysis	129
12.1	MC samples used for the 2005 data analysis	138
12.2	χ^2 values and scale factors from the fit.	156
12.3	Summary of the systematic errors in 2005 data analysis	157
12.4	Differential cross section for b-jets as a function of p_T^{b-jet} in the 2005 data analysis.	158
12.5	Differential cross section for b-jets as a function of η^{b-jet} in the 2005 data analysis.	158
12.6	χ^2 values and scale factors from the fit (extended p_T range). .	158

Chapter 1

Introduction

The study of beauty production is and remains a very interesting topic in high energy physics since it constitutes a rigorous test of perturbative Quantum Chromodynamics (QCD). Given the large mass of the beauty quark which provides a hard scale, the strong force as described within perturbative QCD, should be able to give a precise description of beauty quark production. However, this is not always the case. One of the possible reasons is the presence of more than one hard scale in the process, which leads to some complications in the theoretical calculations (*multiscale problem*). Many measurements¹ of beauty production have been performed by different experiments in high energy interactions, such as UA1 [3], CDF [20, 21] and D0 [31, 30] in $p\bar{p}$ collisions and ZEUS [33] and H1 [14] in ep collisions. Most of the measurements are compatible with the QCD predictions, still characterized by large theoretical uncertainties. Others showed some discrepancies between data and the theoretical calculations, the measured beauty cross sections being higher than the next-to-leading order QCD calculations. Consequently, a better understanding of the beauty quark production mechanism is needed.

One of the main ingredients in the measurement of beauty quark cross-sections is the determination of the beauty fraction in the relevant data sample. Different experimental techniques have been developed to discriminate beauty from charm and light flavour background. These procedures rely often on characteristic features of the beauty production and decay, mostly related to its heavy mass or long lifetime compared to lighter quarks. Well established techniques already used in previous measurements by the ZEUS collaboration make use of the final state muon or electron in semileptonic beauty decays to exploit these features. In fact, the large b mass leads to

¹Quoted are only some of the measurements reviewed in this thesis

high values of the transverse momentum of e.g. the muon relative to the closest jet, p_T^{rel} , making the p_T^{rel} -spectrum of muons(electrons) from beauty decays much harder than the muons(electrons) from a lighter quark. In a similar way, the distribution of the muon impact parameter, δ , can be used statistically to separate the beauty signal from background due to the asymmetry present in muons coming from beauty decay (longer lifetime) compared with the symmetric spectrum of light quarks.

However, all these experimental procedures used so far in ZEUS to measure beauty production are limited to the semileptonic beauty decays. The work presented in this thesis deals with an inclusive beauty measurement in which the hadronic decays are also taken into account. For that purpose, a new method for beauty identification based on the reconstruction of the decay vertices (secondary vertices) has been developed. Due to the relatively long lifetime of b hadrons, beauty secondary vertices are considerably displaced with respect to the interaction vertex compared to the decay vertices of charm and light quarks. This fact will be exploited by means of a signed transverse decay length, L_{xy} , together with the distribution of the invariant mass of the secondary vertices to statistically extract the fraction of beauty in the sample. Beauty cross sections and differential cross sections $d\sigma/dp_T$ and $d\sigma/d\eta$ are then measured in the photoproduction regime. The obtained results are compared to previous measurements and to theoretical predictions from NLO QCD calculations.

A couple of considerations must be pointed out. The method developed here to identify beauty events is only feasible since the new micro-vertex-detector (MVD) was installed during the ZEUS detector upgrade in 2000/2001 and the first physics data were taken in fall 2003. No similar study has been previously done in ZEUS, this work being the first exploratory analysis of this kind. Two analysis are presented here. The first is carried out on the dataset taken during the running period 2004 corresponding to an integrated luminosity of $\mathcal{L} \sim 35 \text{ pb}^{-1}$. During the development of this analysis, several problems were encountered. Their nature relies on problems found of the 2004 dataset such as not enough accuracy in the detector alignment or poor description of main variables by the MC. The problems were partially resolved through the application of different corrections leading to high systematic uncertainties. Nevertheless, the final results are already quite reasonable and the goal to establish a new beauty identification method as well as providing the first inclusive beauty production measurement was reached.

A second, more precise analysis of the 2005 dataset is ongoing. This analysis profits not only from the knowledge gained during the first analysis but also

from an improved detector alignment and tracking software as well as better understanding of the vertexing procedure. The higher statistics and better systematics of this new analysis allows to extend the kinematic range of the measurement to higher transverse momenta of the b -jet. An outlook into this new analysis and the first preliminary results is given at the end of this thesis. Three goals were thus achieved within this work: to develop an inclusive b -tagging algorithm based on the information obtained with the micro-vertex detector, to establish its applicability through a first exploratory analysis and to finally benefit from it by providing improved measurements.

The structure of this thesis is described below:

Chapter 2 gives an overview of the theoretical concepts relevant for lepton-proton interactions, especially those needed for the analysis of beauty photoproduction. The theoretical predictions which will be compared to the obtained measurements are also introduced in this chapter.

Chapter 3 presents some of the most important published results on beauty quark production in high energy collisions, in particular at HERA.

Chapter 4 introduces briefly the HERA collider and in more detail the ZEUS detector, with special emphasis on the description of the Micro Vertex Detector (MVD) which is crucial for the analysis presented in this thesis.

Chapter 5 describes how the physics and detector simulation, based on Monte Carlo (MC) methods, is performed in this work.

Chapter 6 describes the reconstruction of the events, with special attention to the reconstruction of the particle tracks and the final state variables relevant for this analysis.

Chapter 7 presents the algorithms responsible for the primary and secondary vertices reconstruction.

Chapter 8 summarizes the data and MC samples used in this analysis followed by an explanation of the selection applied to them. The secondary vertex b -tagging algorithm is then explained in detail. Finally, the control distributions necessary to investigate the validity of the MC simulation are shown.

Chapter 9 introduces the strategy followed in this thesis to extract of the beauty signal. The corrections applied to the MC simulations in order to achieve a good description of the data are also presented in this chapter.

Chapter 10 presents the b -jet total and differential cross sections, compared to a previous ZEUS result and to NLO QCD predictions. A description of the determination of the systematic errors is given.

Chapter 11 summarizes the measurements obtained so far in this work and presents a b -quark differential cross section as a function of the quark transverse momentum. These results are compared to previous published HERA

results as well as to the theoretical predictions.

Chapter 12 gives an outlook into the analysis of the 2005 dataset presenting the first preliminary results obtained so far. This chapter is structured such that only the differences with respect to the analysis of the 2004 dataset are explained. All procedures and/or reconstruction methods that are common for both analysis are only briefly reminded.

Appendix A gives an overview about the operation of silicon detectors and the criteria that make them specially suitable as particle detectors.

Appendix B gives an outlook in the latest results concerning the analysis of the 2005 dataset.

Chapter 2

Theoretical Overview

In this chapter the theoretical basis relevant for this analysis is presented. After a short introduction into the *Standard Model* of particles physics, the kinematics and properties of ep collisions are described. The chapter then focuses on beauty quark production in the photoproduction regime. The proton structure and its study from HERA measurements is described as well.

2.1 The Standard Model

The *Standard Model* (SM) of particles physics (e.g. [46]) has gradually grown over the last 35 years becoming the most successful attempt to describe nature at small distance scales. It is able to explain most of the known physical phenomena, apart from the gravitational force, currently not included within the Standard Model.

The Standard Model, based on relativistic quantum gauge field theory, describes the electromagnetic, weak and strong interactions: three of the four known fundamental interactions between the elementary particles that constitute the structure of matter. Within the SM, the particles are divided into two groups based on their spin: *fermions* (spin 1/2) and *bosons* (integer spin). The fermions, quarks and leptons, are grouped into three generations, each of them having an anti-particle with the same mass, spin, isospin as the particle, but oppositely signed additive quantum numbers. The interactions between elementary particles are mediated by gauge bosons which can thus be seen as the carriers of the forces between leptons and quarks. Table 2.1 lists the fundamental particles in the Standard Model and their interactions. This dissertation deals with the second heaviest quark, called beauty (or bottom). As it will see along this thesis, its large mass, around 5 GeV, leads to

a special role of this quark within the SM from the experimental and theoretical point of view.

Within the SM, the electromagnetic and weak interactions are unified and explained through the Electroweak theory while Quantum Chromodynamics (QCD) is used to describe the strong interaction. Since the Standard Model is based on a gauge field theory, the three interactions can be described as theories of gauge symmetry i.e. in terms of unitary groups of different dimensions.

The electroweak force is described by a gauge theory based on the $SU(2)_L \otimes U(1)_Y$ symmetry group. $SU(2)_L$ is the group for the weak interactions which only affects left handed leptons and quarks (hence the subscript L), while $U(1)_Y$ is the group of hypercharge. The corresponding gauge bosons are the massless photon of electromagnetism and the massive W and Z bosons of the weak force. The weak gauge bosons get their mass from the spontaneous symmetry breaking introducing a new scalar particle, the Higgs boson, via the so-called Higgs mechanism. Quarks and leptons also acquire mass via interaction with the Higgs boson.

Fermions	Generations			Q_{EM}	Interactions		
	1 st	2 nd	3 rd		EM	WEAK	STRONG
Quarks	$\begin{smallmatrix} u \\ d \end{smallmatrix}$	$\begin{smallmatrix} c \\ s \end{smallmatrix}$	$\begin{smallmatrix} t \\ b \end{smallmatrix}$	$\begin{pmatrix} +2/3 \\ -1/3 \end{pmatrix}$	$\begin{smallmatrix} \checkmark \\ \checkmark \end{smallmatrix}$	$\begin{smallmatrix} \checkmark \\ \checkmark \end{smallmatrix}$	$\begin{smallmatrix} \checkmark \\ \checkmark \end{smallmatrix}$
Leptons	$\begin{smallmatrix} e \\ \nu_e \end{smallmatrix}$	$\begin{smallmatrix} \mu \\ \nu_\mu \end{smallmatrix}$	$\begin{smallmatrix} \tau \\ \nu_\tau \end{smallmatrix}$	$\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}$	$\begin{smallmatrix} \checkmark \\ \times \end{smallmatrix}$	$\begin{smallmatrix} \checkmark \\ \checkmark \end{smallmatrix}$	$\begin{smallmatrix} \times \\ \times \end{smallmatrix}$
Bosons				$(0, \pm 1/0, 0)$	γ	W^\pm/Z^0	g

Table 2.1: The fundamental particles in the Standard Model and their interactions. Each fermion has an associated antiparticle which has opposite electric charge. The value of the electromagnetic charge Q_{EM} is given in multiples of the elementary charge unit.

The strong interaction is described by the $SU(3)$ color symmetry. A quark carries electric charge 2/3 or -1/3 and one of three color charges. The strong interactions are mediated by eight gluons corresponding to the eight generators of $SU(3)$. Gluons also carry colour and thus may also interact among themselves. Leptons are all colorless so that they do not participate in the strong interaction. Since the strength of the strong force increases at low energies or increasing distance scales, no free quarks are ever seen (*confinement*). All observed particles are thus “white” bound states of quarks and/or antiquarks (hadrons).

The strength of the interactions between particles is proportional to the strong coupling constant α_s . QCD is an *asymptotically free* theory i.e. the

value of the coupling constant α_s tends to zero as the energy of the interaction increases or equivalently, at shorter distances. As a consequence of it, at high energies quarks and gluons behave as almost free particles.

2.2 Kinematics of ep interactions

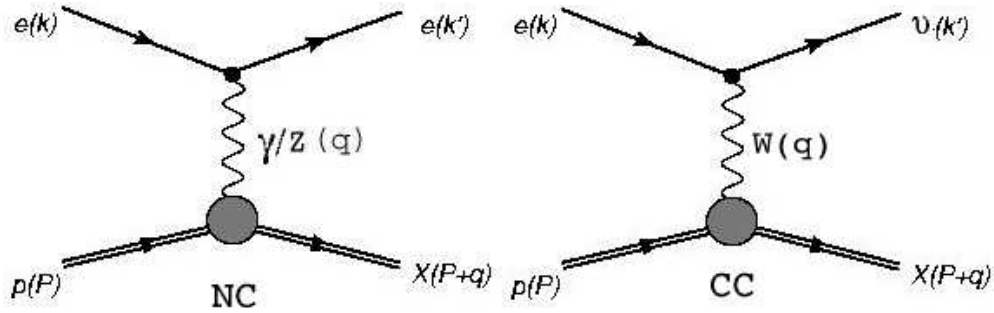


Figure 2.1: Left: Electron-proton scattering in Neutral Current (NC). Right: Electron-proton scattering in Charged Current (CC).

At HERA electrons¹ and protons collide at high energies. Within the Standard Model the interaction between the electron and the proton is characterized by the exchange of a vector boson. If the boson is a photon γ or a Z^0 the process is called *neutral current* scattering (NC), in case the boson is a W^\pm the interaction is called *charged current* scattering (CC). Figure 2.1 shows the generic diagrams of both processes. The kinematics of ep collisions can be explained with the help of this figure as follows. In the diagrams of Fig. 2.1, the incoming electron e^\pm interacts with an incoming proton p with four-momenta k and P respectively. k' denotes the four-momentum of the scattered electron. The exchanged boson has the four-momentum q , given by

$$q = k - k' \quad (2.1)$$

The scattering is described by the following Lorentz invariant variables:

$$s = (k + P)^2 \quad (2.2)$$

$$y = \frac{q \cdot P}{k \cdot P} \quad (2.3)$$

¹Electron (e) denotes both electrons and positrons if not indicated differently

$$Q^2 = -q^2 = -(k - k')^2, \quad 0 \leq Q^2 \leq s. \quad (2.4)$$

$$x = \frac{Q^2}{2P \cdot q}, \quad 0 \leq x \leq 1 \quad (2.5)$$

where Q^2 is the virtuality of the exchanged boson, s is the square of the total center-of-mass energy and y and x are the *Bjorken scaling variables*. The variable y is also called *inelasticity* and describes the relative energy transfer from the lepton to the hadronic system in the proton rest frame.

The Q^2 of a scattering process gives a physical scale to distinguish between two different kinematic regions:

- Photoproduction (PHP or γp) is the regime when $Q^2 \leq 1 \text{ GeV}^2$. Since the propagator of the gauge bosons is proportional to $1/(Q^2 + M^2)$, when the virtuality of the exchanged boson is small, $Q^2 < M_{Z,W}^2$, the exchange of Z/W^\pm bosons is strongly suppressed. In this case, the incoming electron beam can be thought of as a source of quasi-real photons.
- Deep Inelastic Scattering (DIS) is the complementary regime in which events have a high boson virtuality, $Q^2 \geq 1 \text{ GeV}^2$, and large hadronic center-of-mass energy (defined below), $W^2 > m_p^2$.

In the high Q^2 limit and for $E_p \gg m_p$ (*infinite momentum frame*), the Quark Parton Model (QPM) provides the perfect framework for a conceptual understanding of the variable x . The Quark Parton Model is a simple model, previous to QCD, which describes the structure of hadrons as composed of quasi-free pointlike particles, the partons² which were later identified with the quarks. In the infinite momentum frame, the quarks are considered massless particles with negligible transverse momenta so that the ep scattering can be described as the incoherent sum of elastic scattering processes of the lepton off a set of partons. The variable x can be then interpreted as the fraction of the proton momentum carried by the interacting parton.

In case the proton and the electron masses can be neglected, Equations 2.2 and 2.3 can be simplified to:

$$s = (k + P)^2 \simeq 2k \cdot P \quad (2.6)$$

$$y = \frac{q \cdot P}{k \cdot P} \simeq \frac{2P \cdot q}{s}, \quad 0 \leq y \leq 1 \quad (2.7)$$

²Name given to the point-like constituents of the hadrons within the Parton Model proposed by Richard Feynman in 1969

and the virtuality of the exchanged boson Q^2 can be expressed as a function of the Bjorken scaling variables as:

$$Q^2 = sxy \quad (2.8)$$

At HERA the center-of-mass energy is fixed to a value $\sqrt{s} = 318 \text{ GeV}$ (see Chapter 4) so that only two of the four variables remain independent.

The variable W , the invariant mass of the hadronic system recoiling against the scattered lepton, can be expressed as:

$$W^2 = (P + q)^2 = P^2 + q^2 + 2P \cdot q \sim -Q^2 + ys. \quad (2.9)$$

The variable Q^2 can be thought as a “spatial” resolution of HERA to resolve the inner structure of the proton and measure its partonic components using the photon. The quantum wavelength of the boson λ is related to the virtuality of the boson Q as $\lambda \sim h/Q$. HERA can cover a kinematic range up to $Q^2 = 40000 \text{ GeV}^2$ which is then equivalent to a resolution of $\lambda \sim 10^{-18} \text{ m}$ (about one thousandth of the proton size).

The experimental signature used to separate both kinematic regimes is based on the angle of the scattered electron. If $Q^2 \sim 0 \text{ GeV}^2$ i.e. the virtuality of the exchanged photon is small, the electron is scattered at very low angles with respect to the incident direction, escaping into the beam pipe, while if Q^2 is above a few GeV^2 the electron is scattered at a large angle and thus visible in the detector. The measurements presented in this thesis are performed within the photoproduction regimen.

2.3 The perturbative QCD framework

At high energy (or small distances), the strong coupling constant α_s is small enough to allow perturbative theory to be applied (pQCD). Within this approach, the interactions are expressed by perturbative expansions in powers of the coupling constants providing a method for the calculation of the hard parton cross sections and to obtain predictions for the parton dynamics such as the perturbative description of quark-gluon cascades (evolution of the parton density functions).

As an aid to visualize and calculate the interactions between particles in the Standard Model, a graphical method was developed by Richard Feynman in the 1940s and '50s. The Feynman diagrams are used to describe the processes

in a way that provides an intuitive picture of every term in the perturbation expansion of the scattering amplitude for the experiment. A summary of the rules necessary to draw Feynman diagrams can be found e.g. in [41]. Along this chapter, Feynman diagrams will often be used for a better understanding of the physics processed to be studied.

Even in the region where pQCD can be applied, when α_s is small enough, additional difficulties enter in the calculations. Given a certain process, to calculate the cross section all the possible Feynman diagrams for such a process are drawn. Each Feynman diagram represents a term in the series expansion of the scattering amplitude. Feynman diagrams containing loops of virtual particles³ (vacuum fluctuations) lead to divergences in the integration over the full phase space. These divergences are known as *ultraviolet singularities*. Another type of divergences are the so-called *infrared (IR) divergences* which arise from the radiation of partons with very small relative transverse momentum with respect to the original parton (for instance, collinear emission of massless gluons). One way to deal with these IR divergences is to select observables which are insensitive to this particular problem. This type of observable is called infrared-safe.

In general, to get rid of the divergences, a procedure called *renormalization* is applied. Renormalization consists basically in a re-formulation of the problem; the original perturbative expansion is expressed via bare parameters present in the Lagrangian. Those parameters do not coincide with the physical observables (e.g. mass or charge). Therefore, the divergences can be adsorbed through a redefinition of the bare parameters. This requires the introduction of a new dimensional parameter called *renormalization scale* μ_R . All renormalized quantities in the theory, such as the strong coupling constant α_s , will depend explicitly on μ_R . This dependence of the strength of the interaction on the chosen scale is translated into a dependence of α_s with the energy (“running” coupling constant $\alpha_s(Q)$) which is given by the so-called renormalization group equation:

$$\mu^2 \frac{d\alpha_s}{d\mu^2} = \beta(\alpha_s) \quad (2.10)$$

where the β -function is a power serie in α_s . The solution of Eq. 2.10 in the first order of α_s is

³Feynman diagrams represent terms in a quantum calculation which implies that the intermediate stages in any diagram cannot be observed. The particles of such intermediate stages exist only for a limited time and space and can not be detected directly. They are called “virtual particles”. Only the initial and final particles in the diagram represent observable objects, and these are called “real particles”.

$$\alpha_s(\mu^2) = \frac{4\pi}{\beta_1 \ln(\mu^2/\Lambda_{QCD}^2)} \quad (2.11)$$

with

$$\beta_1 = 33 - \frac{2}{3}n_f \quad (2.12)$$

where n_f is the number of quark flavours with a mass lighter than the scale μ . Λ_{QCD} is the scale at which α_s becomes strong as μ^2 is decreased. That means that Λ_{QCD} is the scale that sets the limit on the applicability of perturbative QCD. At or below this scale, the perturbative series in α_s do not converge and pQCD is not useful anymore. Experimentally Λ_{QCD} was found to be of the order of ≈ 200 MeV. This equation also provides evidence for the asymptotic freedom. Due to the logarithmic decrease with the scale μ the value of α_s vanishes for $\mu \rightarrow \infty$ and the quarks are quasi-free.

2.4 Structure of the proton

The Quark Parton Model, as introduced before, assumes that the structure of the proton is made up of quasi-free pointlike particles (partons). In this naive picture, each parton i carries a fraction x of the proton's momentum with the probability density $f_i(x)$. These distribution functions $f_i(x)$, also known as the parton density function (PDF's), are not yet calculable from first principles and have to be determined from the experiments. Both, the NC and the CC cross sections can be expressed in terms of these parton distribution functions. Using them in the regime of medium and low Q^2 where pure γ exchange can be assumed, the cross section can be described as the incoherent sum of elastic scattering processes of the electron off the partons:

$$\frac{d^2\sigma}{dx dQ^2} = \sum_i f_i(x) e_i^2 \cdot \frac{2\pi\alpha^2}{Q^4} \left[1 + \left(1 - \frac{Q^2}{xs} \right)^2 \right] \quad (2.13)$$

A prediction of this model, given the assumption that the proton consists only of the three valence quarks, is the *Bjorken scaling* behavior which consists on the independence of the structure functions (structure of the proton) of Q^2 . The parton model structure functions are given by:

$$F_1(x) = \frac{1}{2} \sum_i e_i^2 f_i(x) \quad (2.14)$$

$$F_2(x) = \sum_i e_i^2 x f_i(x) \quad (2.15)$$

and are interconnected by the Callan-Gross relation:

$$F_2(x) = 2xF_1(x) \quad (2.16)$$

which was experimentally confirmed at SLAC [19] and is a direct consequence of the spin-1/2 nature of the quarks.

This model was however too simple and presented some problems. For instance, if the quarks are the only constituents of the proton, the sum of the momenta of all the quarks should be equal to the momentum of the proton i.e. the momentum fraction should sum to the unity. Experimentally it was found that the fraction of the protons momentum carried by the quarks is approximately 0.5. The QCD improved QPM includes the gluons within the structure of the proton by allowing interactions between the partons via the exchange of gluons. Thus, the proton is seen as a sea of quarks and gluons. The Bjorken scaling behavior was then found to be only an approximate relation. For small values of x , the structure functions increase with Q^2 . This effect is known as *scaling violation*.

Factorization

Factorization consists on the split of the scattering process into two parts: the interaction of high energy partons (hard subprocesses) which can be described within perturbative QCD, and the low energy processes (soft subprocesses) which are not covered by pQCD. According to the factorization theorem [11], the structure function F_2 can be expressed as the convolution of a perturbative part, the *coefficient functions* C_2^i and the non-perturbative parton density functions (PDFs), $f_i(\xi, \mu_f, \mu)$, which as it was seen above, give the probability of finding a parton with a fraction ξ of the proton's longitudinal momentum.

$$F_2 = \sum_{i=\text{parton}} \int_x^1 C_2^i \left(\frac{x}{\xi}, \frac{Q^2}{\mu^2}, \frac{\mu_f^2}{\mu^2}, \alpha_s(\mu) \right) f_i(\xi, \mu_f, \mu) d\xi \quad (2.17)$$

where μ is the QCD renormalization scale and μ_f an additional scale (factorization scale) introduced by the factorization which defines the boundary between the perturbative regime and the non-perturbative regime. The value of this scale depends on the factorization scheme used. The PDFs have to be measured experimentally while the coefficients C_2^i are calculable.

Higher order corrections to the improved QPM, such as softly radiated gluons from the partons, lead to the violation of the Bjorken scaling mentioned before. In fact, if the contributions from the region of collinear gluon or

quark emission are associated with the parton density functions rather than with the hard-scattering processes, the PDFs are no longer independent of Q^2 . The evolution of the PDFs in Q^2 can be then calculated using parton evolution schemes like e.g. DGLAP [4].

2.5 Photoproduction processes and photon structure

The analysis presented in this thesis is carried out within the photoproduction regime. Hence, a detailed description of heavy flavour photoproduction in ep collisions is given here.

Photoproduction processes are of considerable interest in physics since they constitute a transition between DIS and hadroproduction (pp collisions) contributing, among other things, to a better understanding of the proton and the photon structure. If the photon interacts directly with a parton in the proton, the photon structure is not involved in the process and the parton distributions in the proton can be probed down to $x_p \sim 10^{-3}$ for heavy quark production where x_p is the fraction of the proton's momentum entering in the hard process. On the other hand, for $Q^2 \ll 1 \text{ GeV}^2$ the photon may also show a partonic structure, the *resolved photon contributions*, allowing a better study of the structure of the photon.

At HERA the largest contribution to the total ep production cross section is due to photoproduction where, because of the very small photon virtuality $Q^2 \approx 0$, the lepton-proton interaction can be thought of as scattering of a real photon on the proton. For this reason, it is useful to consider the ep scattering at low Q^2 in two parts: the radiation of the photon by the lepton and then the photon interaction with a parton in the proton.

Furthermore, it is well-established that the photon has a partonic structure, having been observed for first time in fixed-target experiments using real photon beams [35]. In this section both parts of the scattering will be treated: first, the rate of the quasi-real photons emitted by the leptons is estimated within the equivalent photon approximation and second, the different behaviors of the photon, based on the finite probability of the process $\gamma \rightarrow q\bar{q}$ to occur, is described. A more complete description can be found e.g. in [62].

2.5.1 The equivalent photon approximation

Electron-proton scattering calculations can be performed within the Equivalent Photon Approximation [37]. Using this approach, the cross section of the ep scattering in the photoproduction regime can be factorized in two terms. Namely, the photon emission from the electron and the γp scattering:

$$\sigma_{hadr}^{\gamma p} = \int dy f_{\gamma,e}(y) \sigma_{\gamma p \rightarrow eX}(y), \quad (2.18)$$

where $\sigma_{\gamma p \rightarrow eX}$ denotes the cross section for the real photon scattering on the proton and $f_{\gamma,e}$ is the flux of quasi-real photon from the electron vertex. The photon flux $f_{\gamma,e}$ can be calculated to high accuracy using the Weizsäcker-Williams approximation⁴ [16]:

$$f_{\gamma,e}(y) = \frac{\alpha}{2\pi} \left[\frac{1 + (1-y)^2}{y} \ln\left(\frac{Q_{max}^2}{Q_{min}^2}\right) - 2m_e^2 y \left(\frac{1}{Q_{min}^2} - \frac{1}{Q_{max}^2} \right) \right], \quad (2.19)$$

where the variable y is the inelasticity defined in Section 2.2, $Q_{min}^2 = m_e^2/(1-y)$ is the kinematic lower limit and $Q_{max}^2 = 2E_e E'_e (1 + \cos \theta_{el})$ the maximum photon virtuality, determined by the experimental geometrical acceptance from the scattered electron's energy E'_e and angle θ_e . α is the fine structure constant.

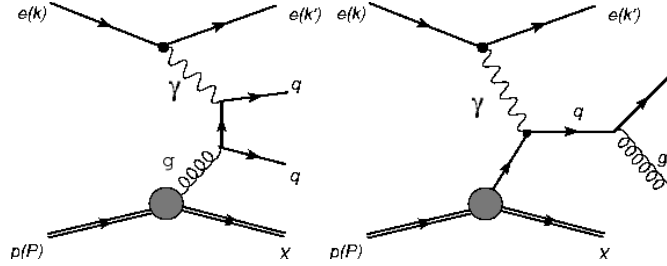


Figure 2.2: Leading order direct photon processes: (Left) Boson-Gluon fusion and (right) QCD Compton

2.5.2 Structure of the photon: direct and resolved processes

As mentioned before, the process $\gamma \rightarrow q\bar{q}$ has a finite probability. The formation of a $q\bar{q}$ pair is only allowed if the time of the $q\bar{q}$ fluctuation lies within

⁴This approximation neglects the virtuality of the photon and the terms involving the longitudinal photon polarization

the life time of the $e\gamma$ state. This requirement is fulfilled at small photon virtualities $Q^2 \ll 1$. According to this, within the photoproduction regime the photon can interact with a parton in the proton in three different ways. If the photon itself couples to the parton in the proton, it is called *direct* photoproduction. On the other hand, if the photon fluctuates in $q\bar{q}$ state before the interaction i.e. acts as a source of partons which undergo hard QCD scattering, the process is called *resolved*. Within resolved processes, the γ fluctuation can be into a bound $q\bar{q}$ state (*hadronic* component) or into an unbound quark pair which is subject to gluon radiation and gluon splitting (*anomalous* component). These two classes of process, direct and resolved, are fully separable only in lowest order (LO) QCD.

The hadronic behavior of the photon (non-calculable perturbatively) has been studied in terms of the *Vector Meson Dominance Model* (VMD) [72], a model which assumes that the virtual photon fluctuates into intermediate $q\bar{q}$ -states (vector meson state) with the same quantum numbers which subsequently may interact with the nucleon. It is a non-perturbative approach and makes use of the structure function of the lowest mass vector meson to estimate the photon structure function F_2^{VMD} .

In the anomalous case, pQCD is used to predict the distribution of partons in the photon. The splitting of a photon into a $q\bar{q}$ pair can be calculated in the QPM. The probability of finding a quark within the photon carrying a momentum fraction x_γ of the initial photon momentum is given by:

$$f_{q/\gamma}(x_\gamma, \mu_\gamma^2) = e_q^2 \cdot \frac{\alpha}{2\pi} (x_\gamma^2 + (1 - x_\gamma)^2) \ln \left(\frac{\mu_\gamma^2 (1 - x_\gamma)}{m_q^2 x_\gamma} \right) \quad (2.20)$$

where e_q is the electric charge of the quark and m_q is the effective mass of the “free” quark. μ_γ is a factorization scale for singularities arising from the electromagnetic vertex (collinear parton emissions). QCD corrections to the simple QPM photon structure function can be calculated from the DGLAP evolution equations. The photon parton densities obey the usual DGLAP evolution equations [4] with the difference that an inhomogeneous term enters in the equation:

$$\frac{df_i^\gamma}{d \log \mu_\gamma^2} = \frac{\alpha_{em}}{2\pi} P_{i\gamma} + \frac{\alpha_s}{2\pi} \sum_j P_{ij} \otimes f_j^\gamma \quad (2.21)$$

where f_i^γ are the partonic densities in the photon and P_{ij} are the parton-parton splitting functions.

Examples of direct photon processes at leading order (LO) are shown in the

figure 2.2. The direct photon process at LO, shown in Fig. 2.2(left), is also called *Boson-Gluon-Fusion* and depicts the dominant process for beauty production at HERA. The diagram can be described as the splitting of a gluon from the partonic substructure of the proton into a $q\bar{q}$ -pair, one of the quarks interacting with the electron by exchange of a photon. Another direct LO process that contributes to the ep cross section is the so-called *QCD Compton Scattering* (see Fig. 2.2(right)) where a quark from the proton interacts with the electron by the exchange of a photon and radiating a gluon after the interaction.

In figure 2.3, resolved photon processes at LO are depicted. In these processes, the photon has a probability to fluctuate into a hadronic state. As a consequence, a gluon (Fig. 2.3(left)) or a quark (Fig. 2.3(right)) of the hadronic structure of the photon can interact with the proton. The latest process, where the photon fluctuates into a pair of heavy quarks and one of these heavy quarks enters the hard interaction, are also commonly referred as to beauty(charm)-*excitation*.

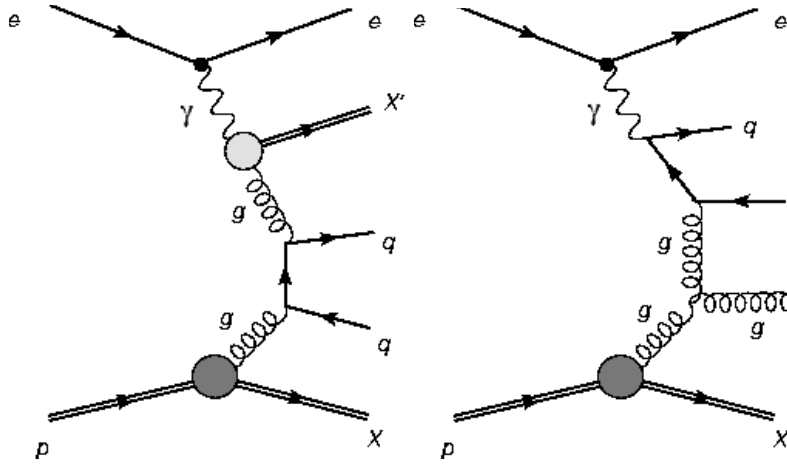


Figure 2.3: Examples of leading order resolved photon processes: (left) hadronic gluon-gluon fusion with interaction of a gluon from the hadronic structure of the photon with the hadronic structure of the proton and (right) hadronic boson-gluon fusion with interaction of a quark from the hadronic structure of the photon with the hadronic structure of the proton.

Summary: Photons

The photon can be described by the superposition of three components:

- The direct component which describes the direct interactions of the photon with the quark of the proton in the hard processes.

- The hadronic component which results from the fluctuation of a photon into a bound $q\bar{q}$ state (hadron). It is described within the vector meson dominance model. This component is very small for heavy quark production.
- The anomalous component which results from the fluctuation of a photon into a quark and anti-quark without the formation of a bound state (flavour excitation). The distributions of the quarks in the photon are predicted by QED and pQCD.

2.6 Parton hadronization

Parton hadronization (or fragmentation) is the process in which, due to QCD color confinement, colorless bound states are formed out of colored partons (quarks and gluons). Thus, right after the production of partons in the ep scattering process, colorless hadrons need to be formed from the initial partons.

Preceding the hadronization, each of these initial partons can at each instant initiate a virtual cascade, branching into a number of additional partons. This perturbative first step is called *parton showering*. The second step is the hadronization process itself, which is a non-perturbative phenomenon and is not yet calculable from first principles. Therefore, some phenomenological models have been developed in order to describe this process and to implement it in the MC models or event generators (see also Chapter 5). The MC event generator used in this thesis, PYTHIA, uses the Lund string model, and thus a more detailed description of this hadronization model will be given here.

2.6.1 Parton showers

Partons (also leptons and photons) can emit other partons, in $a \rightarrow bc$ processes, during the initial and final state (before and after the hard scatter respectively). Each daughter b or c of the process can branch again in turn. This parton emission cascade is also known as *parton shower* (PS) and contributes to higher order corrections not taken into account by the LO hard subprocesses. These higher order effects are important at high energies where they can have a big influence on the topology of the event (e.g. multiple parton emission can give rise to multijet events). The kinematics of each branching is given by two variables: Q^2 which is the transverse momentum

squared of the branching and z defined as the fraction of parent parton energy carried by one of the daughters, being $(1 - z)$ the fraction kept by the other daughter. The probability $P_{a \rightarrow bc}(z)$ to branch is given by the DGLAP equations in the leading approximation of perturbative QCD.

2.6.2 String fragmentation model

The Lund string model [7] is based on the linear confinement picture in which the energy stored in the colour dipole field between a charge and an anti-charge increases linearly with the separation between the charges. The string fragmentation model is very complicated for a generic multiparton state, however the main idea can be explained in terms of the simplest system: a color-singlet $q\bar{q}$ pair, which would be e.g. the final state of an e^-e^+ annihilation event. In this case, as the q and \bar{q} move apart from their common production vertex, a color field, also called *color flux tube* is stretched between the color charge and the anti-color charge with transverse dimensions of the order of typical hadronic sizes ($\sim 1 \text{ fm}$). The tube is assumed to be uniform along its length, which automatically leads to a confinement picture with a linearly rising potential. The more q and \bar{q} separate, the more potential energy is stored until the string may break creating new pairs of $q_i\bar{q}_i$ particles via quantum mechanical tunneling with a probability given by

$$\exp\left(-\frac{\pi m_T^2}{\kappa}\right) = \exp\left(-\frac{\pi m^2}{\kappa}\right) \exp\left(-\frac{\pi p_T^2}{\kappa}\right), \quad (2.22)$$

where m_T , m and p_T are the transverse mass, mass and transverse momentum of the quark respectively, and κ is the string tension i.e. the available energy per unit of length along the color field.

With the creation of these new pairs, the system splits into other colour-singlet systems e.g. $q_i\bar{q}$ and $q\bar{q}_i$. Further breaking can occur if the invariant mass of either of the new two string pieces is large enough i.e. larger than the on-shell mass of hadrons.

The produced new pairs will form hadrons if they carry the correct flavour quantum numbers, otherwise they will just behave as vacuum fluctuations with no observable effects in the experiment. The transverse momentum of the resulting hadron is the sum of the transverse momenta of the constituents quarks. Due to the dependence on the parton mass, the production of strange and, in particular, heavy-quark hadrons is suppressed ($u : d : s : c \equiv 1 : 1 : 0.3 : 10^{-11}$). In fact, heavy quarks are not expected to be produced in the fragmentation process, but only in the hard process or as a part of a parton shower.

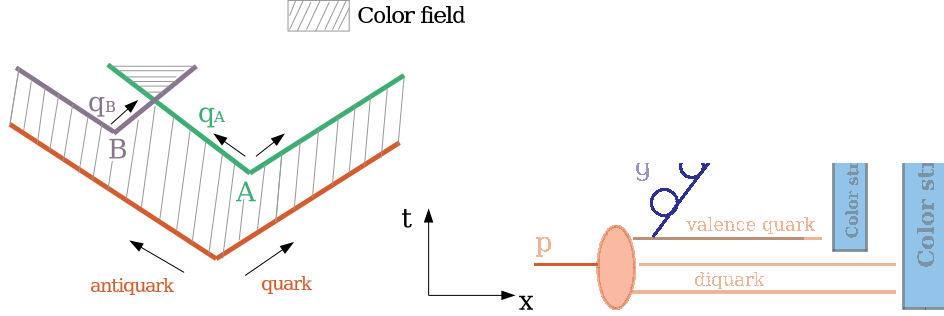


Figure 2.4: Left: Schematic diagram of the space-time development of the $q\bar{q}$ state in breakup situation. The q and \bar{q} move apart as endpoints of a string field until the string breaks up into smaller pieces (producing new $q\bar{q}$ pairs). Right: BGF process at leading order (LO) with color strings. In this process, the proton remnant is the valence quark and valence di-quark. In This example, the color strings are formed between the quark coming from the hard scatter and the valence di-quark of the proton remnant and the antiquark from the hard scatter and the valence quark of the proton remnant.

Figure 2.4(left) shows a schematic representation of the space-time evolution of a $q\bar{q}$ pair. The process of string breakup consists in the production of a new set of $q_i\bar{q}_i$ pairs. The production points of the new pairs, A and B (called vertices), must be spacelike so that the final hadrons have a real mass. There is no interaction between the q and \bar{q} of such a vertex. As illustrated in the figure, the string color field ends on the endpoint charges.

The probability for producing a hadron h with a mass m and a momentum fraction $z = (E + P_{\parallel})_h / (E + P)_q$ where p_{\parallel} is the momentum of the formed hadron h along the direction of the quark q , is given by $f(z)dz$ where:

$$f(z) \sim \frac{1}{z} (1 - z)^a \exp\left(\frac{-bm_{\perp}^2}{z}\right). \quad (2.23)$$

a and b are free parameters which have to be determined from measurements.

As it was discussed above, there is no heavy quark production along the string due to the strong suppression in the tunneling process. However, heavy quark pairs can be, for example, produced at the hard scatter. Due to the mass of the heavy quark, the hadron containing the heavy quark is expected to carry a large fraction of the initial quark momentum so that the fragmentation functions is expected to peak near 1. In this scenario, where the heavy quark is produce at a momentum much larger than its mass, important perturba-

tive effects are expected. The Peterson parametrization [9] is a successful approach (and the one used by PYTHIA) which includes phenomenological non-perturbative effects to the perturbatively calculable result of the heavy quark fragmentation. The Peterson fragmentation function is given by:

$$D(z) \propto \frac{1}{z} \left(1 - \frac{1}{z} - \frac{\epsilon}{1-z}\right)^{-2} \quad (2.24)$$

where the parameter ϵ depends on the heavy flavour considered.

At HERA, the final state is more complicated than a color-singlet $q\bar{q}$ pair and thus the fragmentation model. In ep interactions, the dominant process at LO is the Boson Gluon Fusion. In this case, as depicted in Fig. 2.4(right), a color-octet gluon is taken out from the proton leaving behind a quark and a di-quark as proton remnant. The quark from the hard scatter forms a color string with the di-quark while the anti-quark from the hard scatter is connected via another string to the remnant valence quark.

2.7 Beauty production and decays

The charm, beauty and top quarks are generally called *heavy quarks*, since their mass are larger than that of the *light quarks* (u,d,s) and than the energy scale at which the quarks are confined into hadrons, Λ_{QCD} . Because of their large mass, a reliable description by perturbative QCD should be possible since the heavy quarks masses provide a natural infrared cut-off for the theoretical calculations so that a finite order-by-order perturbative expansion in α_s can be obtained. This is particularly useful since it assures at least one hard scale in all processes where heavy quarks production is studied. Thus, heavy flavour production is an interest test of the theory of the strong interactions QCD.

The lowest order process of heavy quark production is the BGF (Fig. 2.2 left). In a BGF process, a heavy $Q\bar{Q}$ pair can be produced if the squared center-of-mass energy of the γ^*g system, \hat{s} , is:

$$\hat{s} = (\gamma^* + g)^2 = (q + \xi P)^2 > (2m_Q)^2, \quad (2.25)$$

where m_Q is the mass of the heavy quark⁵ and γ and $g = \xi P$ are the photon and gluon four-momenta respectively. Two kinematic regions can be distinguished for heavy quark production. In the region $Q^2 \leq (2m_Q)^2$, BGF is the lowest order production process for a $Q\bar{Q}$ pair of mass $2m_Q$. For high $Q^2 \gg (2m_Q)^2$ the splitting of a gluon into a virtual $q\bar{q}$ - pair can be seen as

⁵capital “Q” here denotes a heavy quark

occurred inside the proton and the QPM is applicable for the production mechanism.

2.7.1 BGF cross section

The LO cross section for heavy $b\bar{b}$ -pair production in BGF can be calculated [56] as:

$$\hat{\sigma}_{BGF} = \frac{\pi e_b^2 \alpha \alpha_s}{\hat{s}} \left[(2 + 2\omega - \omega^2) \ln \frac{1 + \chi}{1 - \chi} - 2\chi(1 + \chi) \right] \quad (2.26)$$

where $\hat{s} = (p_b + p_{\bar{b}})^2$ is the squared centre-of-mass energy of the $b\bar{b}$ pair and e_b the electromagnetic beauty quark charge. The variables ω and χ are defined as $\omega = 4m_b^2/\hat{s}$ and $\chi = \sqrt{1 - \omega}$. Due to the lower charge of the beauty quark and its higher mass, the factor e_b^2/\hat{s} is smaller for beauty than for charm BGF processes, leading to a strong suppression of the beauty cross section with respect to the charm cross section. Other contributions to the beauty production cross section will be discussed in section 2.8.2.

2.7.2 B -hadrons and decays

B -hadrons are produced as result of the hadronization process of b quarks. Since the lifetime of the B -hadrons is around 1 ps, they can only be detected through their decay particles which reach the detector. In this section, the properties of the B -hadrons and their decays are outlined.

Since the probability for $q\bar{q}$ creation from the vacuum depends on the $q\bar{q}$ mass, the most common B hadrons are B^\pm and B^0/\bar{B}^0 , followed by B_s^0 and the Λ_b^0 baryons. Table 2.2 summarizes some of the properties of the B -hadrons [34].

The simplest picture for the hadron decay is given by the *spectator model*. In this model, the decay process of the B -hadron is interpreted as the decay of a free beauty quark via the charged current weak interaction. The light quarks are assumed not to participate in the dynamics (hence, the model name “spectator”). Thus, QCD effects such as binding effects and gluon radiation are neglected. These assumptions are justified by the large mass of the b quark compared to the other quarks in the hadron (mainly u or d) which leads to a suppression of the QCD effects. B -hadron decays occur predominantly through the Cabibbo favored $b \rightarrow c$ transition, where the beauty quark decays into a virtual W and charm quark. All B mesons decay that do not occur through it are known as rare B decays.

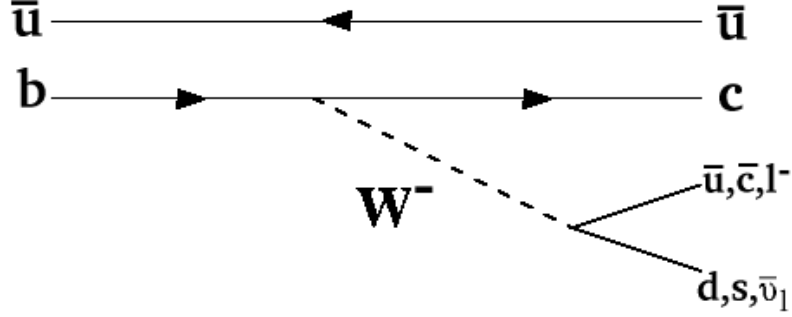


Figure 2.5: Diagram for B^- decays in the spectator model.

The B -hadron decays are classified in semileptonic or hadronic depending on the W decay, which can go respectively to a charged lepton and its associated neutrino, or a $q\bar{q}$ pair. In Fig. 2.5 a few examples for B^- decay within the spectator model are showed. Depending on the W^- decay, the final states depicted are semileptonic, $D^0 l \bar{\nu}_l$, where l can be an e^- , μ^- or τ^- , or hadronic, $D^0 \pi^-$, $D^0 D_s^-$, $D^0 K^-$, $J/\phi K^-$ and $J/\phi \pi^-$.

Hadron	quark content	mass/[MeV]	decay length $c\tau$ /[μm]
$B^+(B^-)$	$ub(\bar{u}b)$	5279.0 ± 0.5	501
$B^0(B^0)$	$db(\bar{d}b)$	5279.4 ± 0.5	460
$B_s^0(B_s^0)$	$sb(\bar{s}n)$	5369.6 ± 2.4	438
$\Lambda_b(\Lambda_b)$	$udb(\bar{u}\bar{d}b)$	5324 ± 9	368

Table 2.2: Examples of B -hadrons [34]

2.8 Beauty production models and predictions

In this work, data distributions are compared to MC simulations (LO+PS) and measured cross sections are compared to the NLO QCD predictions. In the following the event generator as well as the NLO calculations used in this thesis are described.

2.8.1 The PYTHIA MC event generator

PYTHIA is the general-purpose event generator adopted to generate the MC samples used in this thesis. PYTHIA can simulate e^+e^- , pp and ep collisions. It is based on leading order calculations of the matrix elements (ME) of the hard subprocess (indicated as ME in Fig. 2.6) supplemented with parton

showers in the initial and final state (indicated as PS in Fig. 2.6). The soft contributions (low energy processes) are simulated using phenomenological approaches. Since the parton showers effectively approximate higher order effects, MC generators of this kind are commonly referred to as LO+PS generators. For a complete description of it, see [36].

The initial input of PYTHIA are the parton distributions of the colliding particles, at HERA leptons and protons. In particular, for photoproduction events the Weizsäcker Williams Approximation was used to generate the spectrum of photons radiated from the incoming lepton. Since the philosophy of PYTHIA is to describe the hadronization processes in as much detail as possible, special attention is given to the process of parton fragmentation. PYTHIA fragmentation is based on the string fragmentation or Lund model which was explained in section 2.6.2. The hadrons produced in the fragmentation process, in secondary decays or as a part of the beam remnant treatment, may be stable or unstable. The subsequent decay of the unstable heavy hadrons is done according to the branching ratios (BR) and decay spectra measured experimentally by other experiments. These measurements reflect general properties of B-hadrons not restricted to the special production environment. In figure 2.7, the electron energy spectrum in the B-hadron center-of-mass system as implemented in PYTHIA are compared to Belle and Babar data [39]. Good agreements is found. It is then reasonable to expect a good description by the MC models of the B-hadron distributions at HERA. The BR of all leptonic channels and all hadronic two-body decay channels are explicitly listed, while hadronic channels with three or more particles are only given in terms of a quark content.

2.8.2 Next-to-leading order processes

As stated before, heavy flavour production is considered to be reliably calculable in perturbative QCD since the mass of the heavy quark (here the b quark mass) provides a hard scale so that α_s is sufficiently low to apply pQCD. Perturbative QCD calculations of heavy quark production are available up to next-to-leading order (NLO). However, the hard scale set by the heavy quark mass can lead to some complications in the theoretical calculations since in many physical processes more than one hard scale might be present, like e.g., the transverse momentum of the quarks p_T or the virtuality of the photon Q^2 . Since the perturbative expansion can not be optimized for all the scales at one, different calculation approaches have been developed assuming one hard scale in each. The discussion of the different approaches is restricted to the low Q^2 region (PHP).

- **massive scheme:** In the massive scheme, the quarks are treated as heavy particles. Only light quarks and gluons are considered to be active flavours in the initial state while the heavy quarks are produced dynamically in the hard scattering. The heavy quark mass is fully taken into account as a parameter. It acts as a cut-off for initial and final state collinear singularities and sets the scale for the perturbative calculations. Nevertheless, in the NLO perturbation series there are terms $\propto \alpha_s \ln(p_T^2/m_q^2)$ that can be neglected when $P_T \sim m_q$ but which become large in the high- p_T limit. These terms arise e.g from collinear gluon emissions by the heavy quarks at large transverse momenta and gluon or photon splitting into a heavy $q\bar{q}$ -pair. For this reason, this approach is reliable in the phase space region where the transverse momentum P_T of the heavy quark is similar to its mass. This scheme is also known as *Fixed Order* (FO) scheme.
- **massless scheme:** In this approach, the heavy quarks are treated as additional active flavours in the proton and photon together with the light quarks and can appear in the final state through heavy quark excitation processes. The terms proportional to $\ln(p_T^2/m_q^2)$, which become large in the high- P_T limit, are controlled in the so called *resummed calculations* to next-to-leading logarithms (NLL). These calculations resums the $\alpha_s [\ln(p_T^2/m_q^2)]^n$ terms to all orders assuming that all the quarks are massless (heavy quark masses are set to zero) and introducing a fragmentation function to absorb the final state divergences. The predictions of this scheme should be valid for transverse momenta much larger than the heavy quark mass.

There is also an approach, known as FONLL [57], which combines a massless resummed calculation with a massive fixed order one .

Given the kinematical range of this analysis, the obtained measurements will be compared to the NLO calculations in the massive scheme using the program FMNR as described in the following subsection. In Figures 2.8 and 2.9, Feynman graphs of real and virtual corrections, contributing to the NLO cross sections, are shown. Comparing to LO diagrams (BGF in massive scheme, Fig 2.2 left), the main difference rely on additional radiation of hard gluons and the interference with virtual corrections.

The NLO QCD prediction

The cross sections measured in this thesis will be compared to NLO QCD predictions based on the program FMNR by Frixione *etal.* [73]. FMNR provides parton level events with two or three outgoing partons and calculates

photoproduction cross sections. The calculation is implemented in the massive scheme i.e. only light quarks (u,d,s) and gluons are active partons in the initial state while charm and beauty quarks are generated dynamically in the hard process via boson-gluon fusion diagrams (see above).

Jets are found applying the k_T -Algorithm (see Sec. 6.3) to the partonic final state generated by the program. The fragmentation of the $b(\bar{b})$ quarks into B hadrons is done according to the Peterson fragmentation function.

The following parameter set is used for the calculation:

- Photon PDF: CTEQ5ML [24],
- Photon PDF: GRV-HO [58],
- beauty quark mass : $m_b = 4.75 \text{ GeV}$,
- factorization and renormalization scales are set to the transverse mass of the beauty quark pair as it is proposed by the authors for the $b\bar{b}$ cross sections:

$$\mu_R = \mu_F = \langle m_T \rangle = \sqrt{\frac{1}{2}((P_T^b)^2 + (P_T^{\bar{b}})^2) + m_T^2}$$

where $p_T^{b\bar{b}}$ is the transverse momentum of the $b(\bar{b})$ -quark in the laboratory frame,

- $\Lambda_{QCD} = 0.226 \text{ GeV}$ according to the value of α_s used in the CTEQ5M PDF,
- Peterson Fragmentation with $\epsilon = 0.0035$

In order to allow a direct comparison with the measured hadron-level cross sections of this thesis, the NLO cross sections, calculated for jets made of partons, were corrected for jet hadronization effects using the PYTHIA Monte Carlo event generator as explained in [86]. In the analysis presented here, where dijets event are selected satisfying the cuts described in Sec. 8.2.2, the correction factor in the i^{th} bin are defined as $C_{had} = N_i^{had}/N_i^{part}$, where $N_i^{had}(N_i^{part})$ is the number of events in the given kinematic range having two hadronic (partonic) jets in the final state passing the same requirements than in the analysis. These factors are evaluated comparing massive (massless) hadronic jets to massive(massless) partonic jets . The hadronic jets were reconstructed in different ways: using the final state particles ,including the b decay products, or all the particles before the beauty decay, including the initial B meson.

The main systematic uncertainties on the NLO calculations come from the

b -quark mass and the renormalization and factorization scales. The uncertainties are obtained by varying m_b and $\mu_R = \mu_F$ simultaneously from $m_b = 4.5 \text{ GeV}$ and $\mu_R = \mu_F = m_T/2$ to $m_b = 5.0 \text{ GeV}$ and $\mu_R = \mu_F = 2m_T$. Also the parameters Λ_{QCD} , ϵ and proton PDF are varied in order to determine their influence on the cross section calculations. However, when adding the errors in quadrature, their contributions to the final error are negligible. A more detailed description of how the uncertainties are calculated can be found in [33].

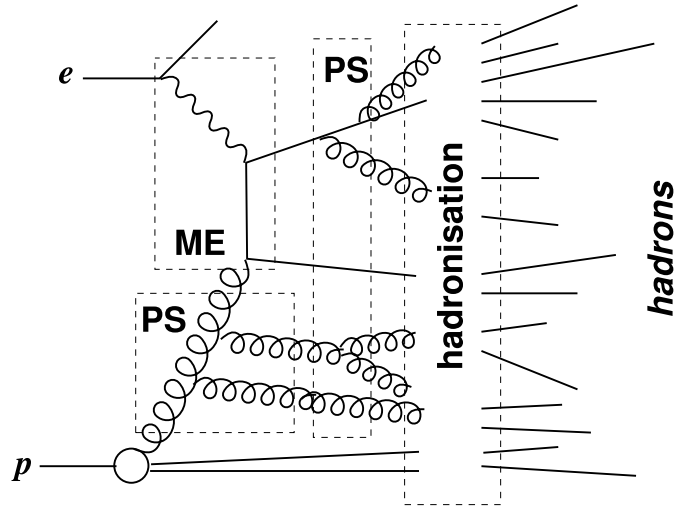


Figure 2.6: Schematic structure of ep generator in a BGF event. ME and PS are the abbreviation for “matrix elements” and “parton showers” respectively.

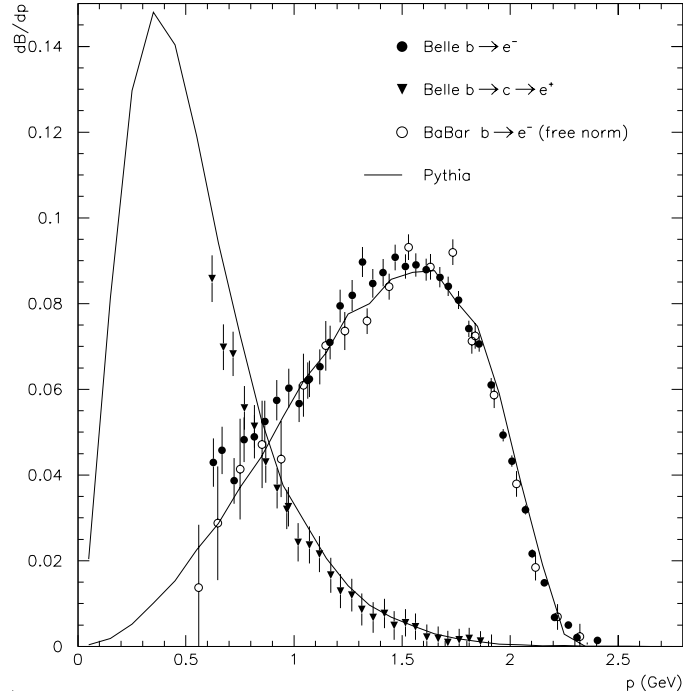


Figure 2.7: Differential branching ratio dB/dp for the electron momentum in the B-hadron centre-of-mass system as measured by Belle for direct (circles) and indirect (triangles) B-meson decays compared to the PYTHIA prediction (line). The momentum distribution measured by BaBar is shown with arbitrary normalization (empty circles).

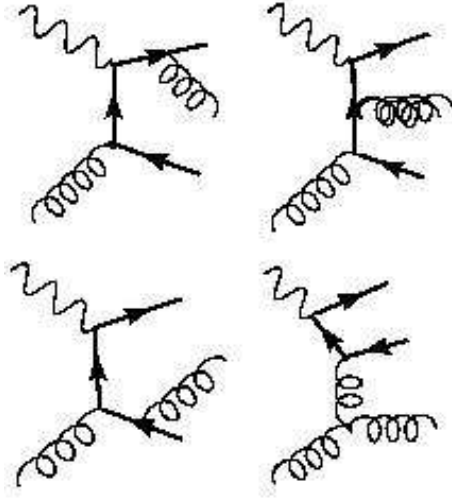


Figure 2.8: Real NLO QCD contributions (gluon radiation) to heavy quark production in the massive scheme.

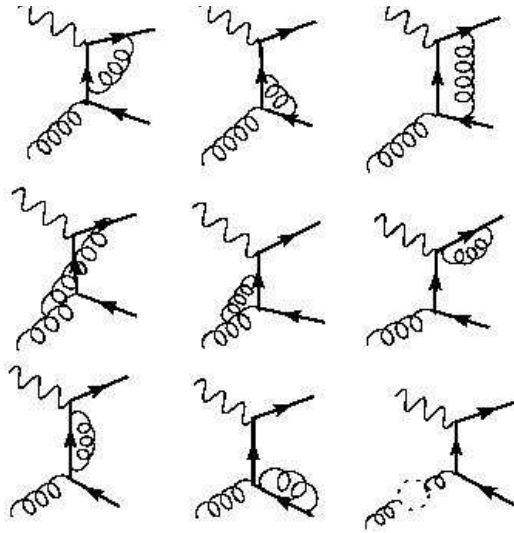


Figure 2.9: Virtual NLO QCD contributions to heavy quark production in the massive scheme.

Chapter 3

Beauty Production - An experimental Overview

Since the discovery of the beauty quark, also called bottom, in 1977 at Fermilab [40, 85], beauty production has been intensively studied since it plays an important role from the theoretical and experimental points of view. This chapter has the goal of giving a collection of some of the most important beauty cross section measurements obtained by different collaborations from the first fixed-target experiments to the colliders HERA and Tevatron.

3.1 Beauty production: from fixed target to collider experiments

The first evidence of the beauty quark was provided in 1977 by a team of physicists led by Leon Lederman [85] working on the experiment E288, a fixed-target experiment studying the results of 400 GeV proton-nucleus collisions at Fermilab. This evidence was a direct consequence of the observation of a dimuon resonance at 9.5 GeV, the Υ meson, a bound state of a beauty quark and its antiquark as confirmed soon after its discovery (see Fig. 3.1). The $\Upsilon = b\bar{b}$ mesons are also known as *hidden* beauty mesons, characterised by having the beauty flavour quantum number $B = 0$. On the other hand, when $B \neq 0$ the term *open* beauty mesons is used. The bound $b\bar{b}$ states decay strongly, with b and \bar{b} annihilating each other. This process is suppressed by the OZI-rule [78].

Many measurements of open beauty production cross sections followed after the discovery of the beauty quark. Nearly all experiments¹ reported cross

¹The exception is the single B^+ meson cross section by CDF in Run I

sections for a global mixture of all different beauty hadrons. Fixed-target experiments were carried out with different type of beam particles of varying energies and many different targets contributed to important results on the basic properties of heavy quark production in strong interactions. However, such measurements at fixed-target experiments were (and remain) very demanding due to the minimum energy required for beauty production which is barely reached by the fixed-target experiments. The beauty cross section is measured at much higher center-of-mass energy by the UA1 and the CDF collider experiments at the $S\bar{p}\bar{p}S$ collider and Tevatron respectively. The analyses used semi-leptonic beauty decays to muons to identify beauty quarks. Table 3.1 summarizes some of the results obtained from fixed-target experiments and collider experiments.

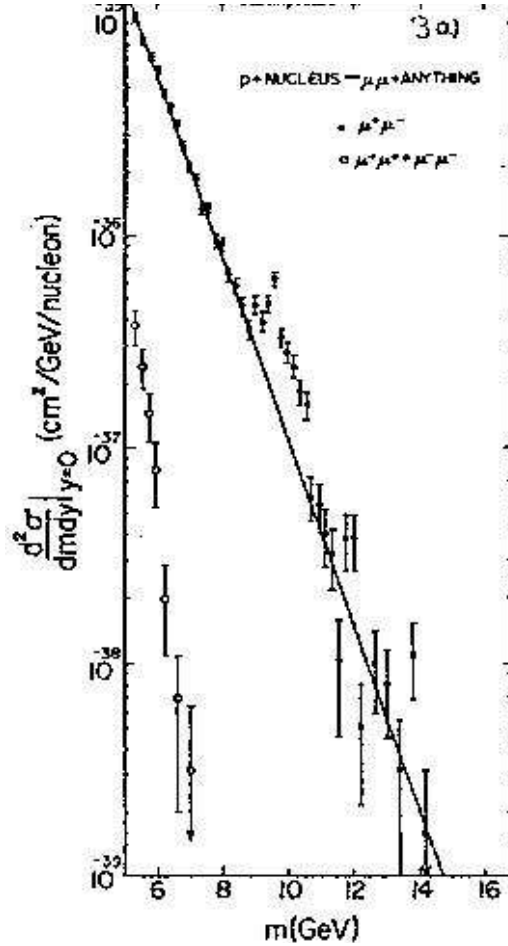


Figure 3.1: Observation of dimuon resonance at 9.5 GeV: the Υ meson. The figure shows the measured dimuon production cross section as a function of the invariant mass of the dimuon pair.

Experiment	Beam- E_{lab} [GeV]	Target	Phase space window	$b\bar{b}$ events
NA10 [27]	π^- -286	W	$x_F > 0$	14
E672/E706 [29]	π^- -515	Be	$x_F > 0$	8 ± 3
E789 [51]	p -800	Au	$0 < x_F^{J/\psi}, p_T^{J/\psi} < 2$	19 ± 5
HERA-B [2]	p -920	C,Ti,W	$-0.35 < x_F^{J/\psi} < 0.15$	83 ± 12
UA1 [3]	$p\bar{p}$ at $\sqrt{s} = 630$ GeV		$ y < 1.5, p_T^\mu > 6$	2859
CDF Run I [20]	$p\bar{p}$ at $\sqrt{s} = 1.8$ TeV		$ y < 1.0, p_T^\mu > 6$	387 ± 32
CDF Run II [21]	$p\bar{p}$ at $\sqrt{s} = 1.96$ TeV		$ y < 0.6$	38078

Table 3.1: Collection of beauty production measurements at fixed target experiments and collider experiments $p\bar{p}$.

In figure 3.2 some of these measurements are compared to the LO+PS calculations performed with PYTHIA (see Sec. 2.8). In Fig. 3.2 (a) where results from pion collisions are shown, the prediction is scaled with empirical factors $k = 1$, $k = 2$ and $k = 3$, while in figure 3.2 (b) the results are based on proton collisions with different targets in fixed-target experiments and $p\bar{p}$ data from UA1 and CDF. The PYTHIA prediction was arbitrarily scaled up with a factor $k = 2$ and refers to pp collisions. This prediction can be compared to the UA1 and CDF data points since the beauty production cross sections at such high energy is dominated by gluon-gluon fusion processes, being pp and $p\bar{p}$ predictions identical.

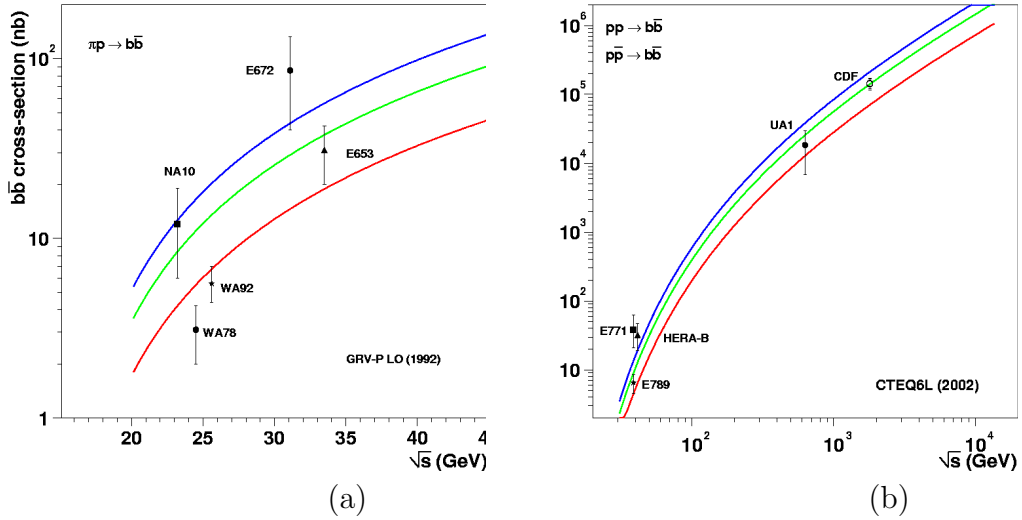


Figure 3.2: Beauty production cross sections in (a) πp and (b) $p\bar{p}$ collisions. In figure (a), PYTHIA predictions are scaled up with a factor $k = 1, 2, 3$. In figure (b), the measurements are compared to predictions from PYTHIA for $pp \rightarrow b\bar{b}$ (see text) scaled up with a factor $k = 2$.

3.2 Beauty quark production at Tevatron

The experiments $D\bar{0}$ and CDF at the Tevatron $p\bar{p}$ collider at Fermilab² provide beauty measurements at the highest center-of-mass energy reached so far. Table 3.2 summarizes the energies at the different data taking periods.

Period	Time	Center-of-mass Energy
Run Ia	1992-1993	$\sqrt{s} = 1.8 \text{ TeV}$
Run Ib	1993-1995	$\sqrt{s} = 1.8 \text{ TeV}$
Run IIa	since 2001	$\sqrt{s} = 1.96 \text{ TeV}$

Table 3.2: Tevatron running periods.

Early cross section measurements at CDF [23] and $D\bar{0}$ [30] using Run Ia data reported compatibility of the NLO QCD shape but a normalization difference of a factor ~ 2 . Other results from the $D\bar{0}$ collaboration [31] showed a good agreement between the data and the NLO calculations (Fig. 3.3(a)).

Recent improvements both in experiment and theory led to solve the observed discrepancies [22]. Figures 3.3(b) shows the CDF preliminary results of Run II data. The measurements are compared to refined NLO calculations, including resummations of logarithmic corrections for all perturbative order and proper treatment of hadronization. The normalization difference seems to be solved by these improved NLO calculations.

Special attention is given here to one of the latest measurements on beauty production from CDF carried out using a small sample of data at $\sqrt{s} = 1.96 \text{ TeV}$ [54]. The method for beauty quark identification used in this analysis is similar to the one used in this thesis. Taking advantage of the long lifetime of the b hadrons, the identification of the b -hadrons is based on the significance of the impact parameter and of the decay length L_{xy} . Figure 3.4(a) shows the distribution of the mass of the secondary vertex for jets with $82 < p_T^{jet} < 90 \text{ GeV}$. As expected, the b -jet spectrum is much harder than for charm and light flavours jets. Figure 3.4(b) shows a preliminary measurement of the b -jet differential production cross section as a function of p_T^{jet} . It is compared to the predictions of the PYTHIA Monte Carlo and a reasonable agreement is found considering the fact that PYTHIA calculations are at LO only. A comparison to NLO pQCD is in process.

²Fermi National Accelerator Laboratory

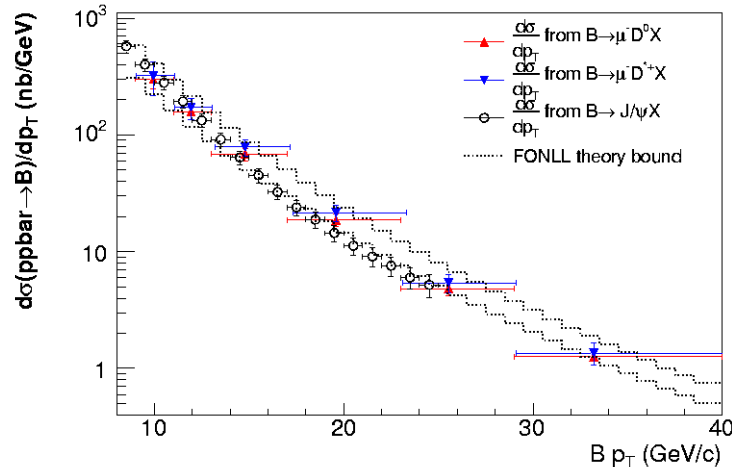
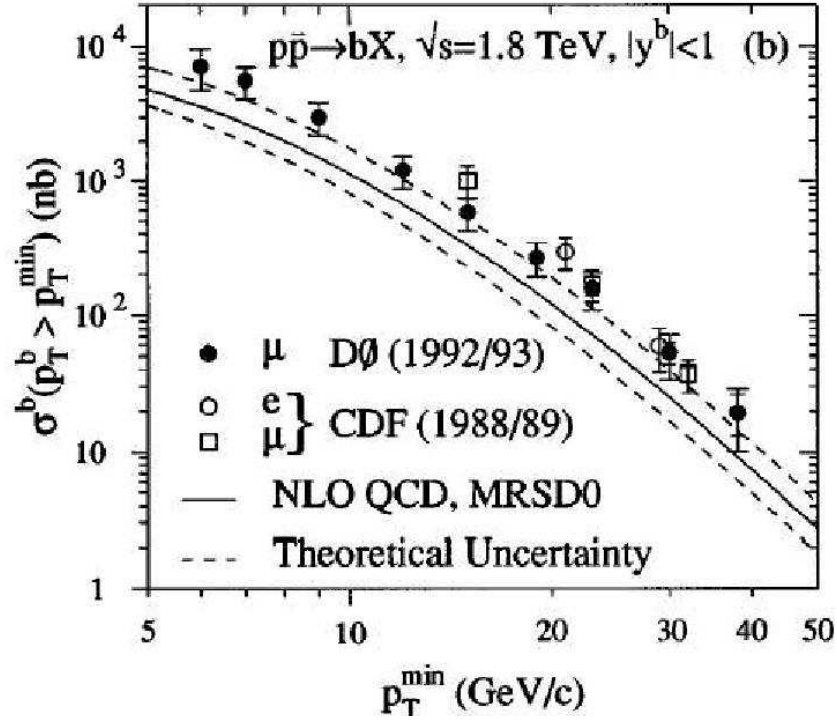


Figure 3.3: (Top) Measurements of beauty quark production cross section for $|y|^b < 1.0$ from $D0$ and CDF compared to the NLO QCD prediction and (Bottom) from CDF Run II in the rapidity range $|y| < 0.6$ with FONLL theory bounds (see text).

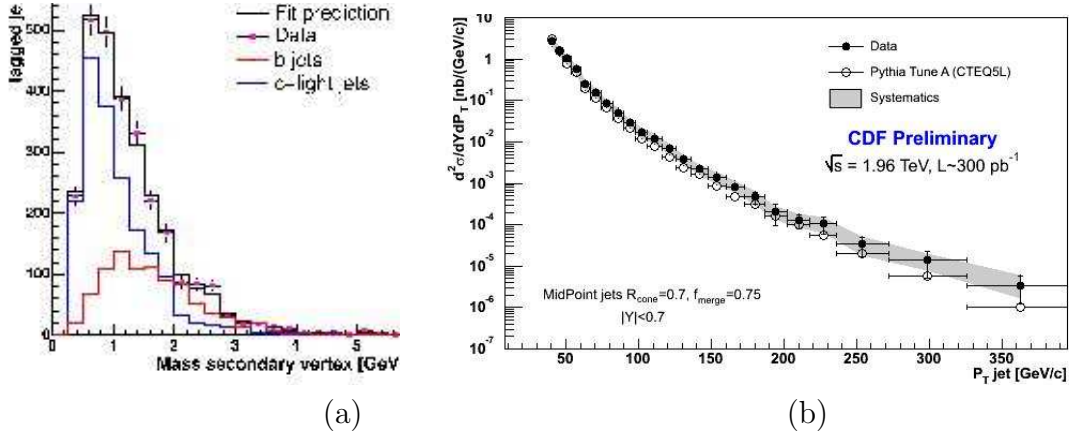


Figure 3.4: (a) Mass of secondary vertex distribution for jets with $82 < p_T^{jet} < 90 \text{ GeV}$. (b) Differential inclusive b -jet production cross section as a function of p_T^{jet} . The measurements are compared to the predictions of PYTHIA. Data points include only statistical errors.

3.3 Beauty production at HERA

Recent measurements of beauty production at the lepton-proton collider HERA, working at a center-of-mass energy of 318 GeV, are presented here, with special emphasis on two results which are of special interest for this thesis. First, the measurement from the ZEUS experiment of beauty production in the photoproduction regime in dijet plus muon events [33] and second, a measurement of beauty dijet cross sections in the photoproduction regime from the H1 experiment [14]. While the first provided the starting point for the beauty quark measurements in this thesis, the interest of the second one relies on its beauty identification method which is very similar to the one applied in this thesis (never used in ZEUS before).

3.3.1 ZEUS beauty photoproduction measurements in dijet+ μ events

This section briefly describes one of the published results from the ZEUS Collaboration on beauty production with semi-leptonic decay into muons [33]. First μ -jet cross sections are measured and then corrected to obtain the cross sections for b -jets in dijet events: $\sigma(ep \rightarrow e'jjX)$. These measurements motivated the study presented in this thesis providing the starting point of the analysis.

The beauty quarks are identified using the p_T^{rel} method. This method exploits

the transverse momentum of the muon candidate with respect to the direction of the associated jet. For beauty quarks, due to the large b -mass, the p_T^{rel} distribution is harder than for charm and lighter flavours. This allows to fit the light quark, charm and beauty quark Montecarlo distributions to the measured p_T^{rel} distribution in order to extract the fraction of muons coming from beauty quark decays in the data.

The μ -jet cross-sections, obtained by counting muons (or b -jets), are corrected to obtain the cross-sections for b -jets using PYTHIA and accounts for the $b \rightarrow \mu$ branching ratio, including direct semileptonic decays ($b \rightarrow \mu X$) and indirect decays into muons³. A b -jet is defined as a jet containing a B (or \bar{B}) hadron. Figure 3.5 shows the differential cross sections as a function of the pseudorapidity, $d\sigma/d\eta^{jet}$, and the transverse momentum, $d\sigma/dp_T^{jet}$, of a b -jet.

The results are compared to NLO QCD predictions from FMNR [73] (see chapter 10) and to the PYTHIA MC model. The prediction is in good agreement with the data in both differential distributions.

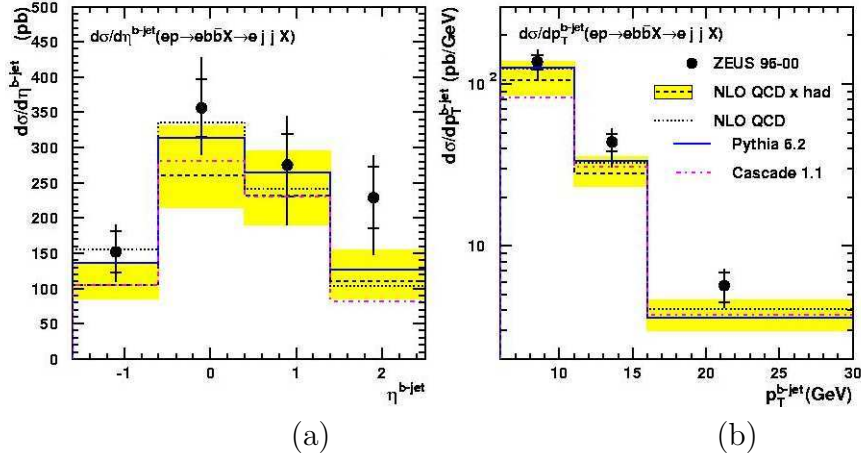


Figure 3.5: Differential beauty photoproduction cross sections as a function of a) the pseudorapidity $d\sigma/d\eta^{jet}$ and b) the transverse momentum $d\sigma/dp_T^{jet}$ of a jet containing a B hadron for the process $ep \rightarrow eb\bar{b}X \rightarrow ejjX$. The inner error bars show the statistical errors while the outer error bars indicate the statistical and systematic uncertainties added in quadrature. The measurements are compared to the predictions of the PYTHIA Monte Carlo model and the NLO QCD calculations corrected for hadronization effects. The shaded band shows the estimated uncertainty on the theoretical prediction.

³Indirect cascade decays into muons via charm, anti-charm, τ^\pm and J/ψ .

3.3.2 H1 beauty photoproduction measurements in di-jets events

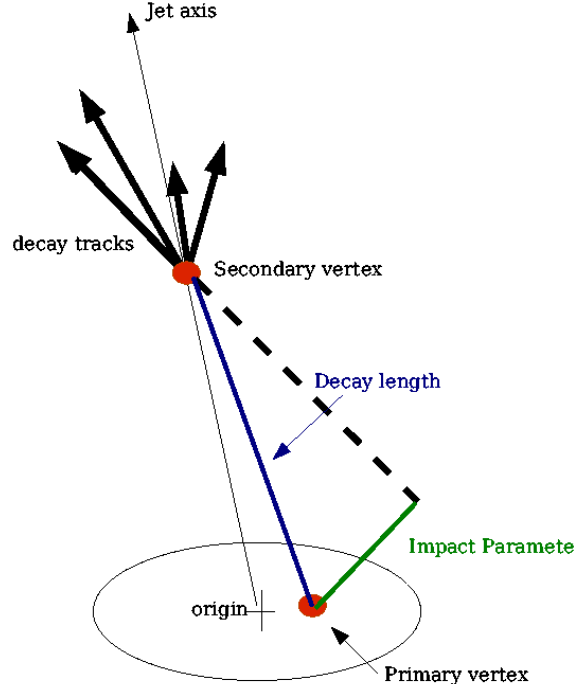


Figure 3.6: Schematic illustration of the impact parameter in the $r\phi$ -plane.

The H1 collaboration provides similar measurements to the results from ZEUS shown above but replacing the p_T^{rel} method in semi-leptonic decays by an inclusive analysis using a method based on the impact parameter (see Fig. 3.6) the beauty identification as described below.

Events are selected with at least two jets of transverse momentum $p_T^{jet} > 11(8)$ GeV. The impact parameter δ of selected tracks is calculated as the transverse distance of closest approach to the reconstructed primary vertex (event vertex). The impact parameter is given a sign which is positive if the angle between the axis of the associated jet and the line joining the primary vertex to the point of closest approach of the track is less than 90° , and is negative otherwise. The distribution of δ is shown in figure 3.7(a). Decays of long-lived particles are characterized by positive impact parameters, whereas the finite track resolution yields a symmetric distribution centered on zero for short-lived particles. Figure 3.7(b) shows the impact parameter significance, defined as $\delta/\sigma(\delta)$ where $\sigma(\delta)$ is the error on δ . The distribution of S_2 (Fig. 3.7(c)) show the significance of the track with the second highest

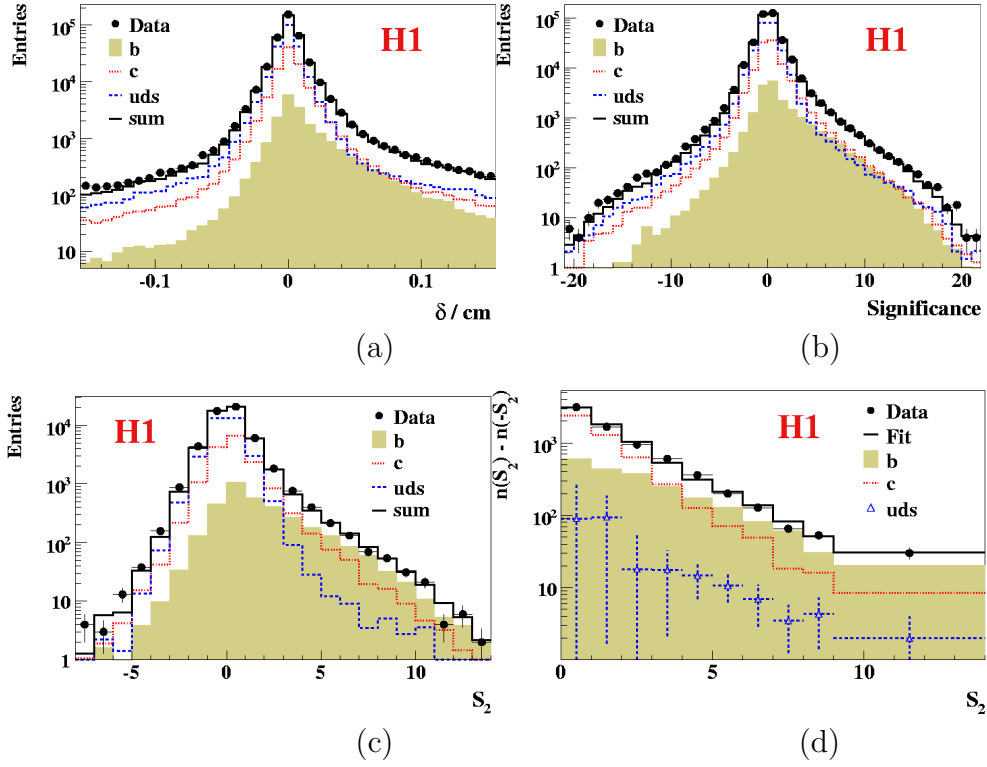


Figure 3.7: Distributions of a) the signed impact parameter σ of selected tracks, b) the signed significance, c) the significance S_2 of the track with the second highest significance in jets with two or more selected tracks and d) the subtracted signed significance for the sample with two or more tracks.

absolute significance and same sign of δ in jets with two or more selected tracks.

In order to reduce systematic errors due to the resolution of δ and the light quark contribution, the negative bins in S_2 are subtracted from the positive ones (Fig. 3.7(d)). The beauty fraction in the data is extracted using a simultaneous least squares fit of the Monte Carlo simulated c and b distributions to the measured $S_2^+ - S_2^-$ and $S_1^+ - S_1^-$.

The measured differential cross sections as function of p_T^{jet} and η^{jet} are shown in figure 3.8 (a) and (b) respectively. Measurements are compared to PYTHIA predictions and NLO QCD calculations. The comparisons show that the data distributions are consistent both in shape and normalization with the calculations in pQCD to next-to-leading order. A tendency of the data to lie above the NLO predictions is observed, specially for the beauty differential cross sections as a function of the transverse momentum of the jet for lower values of p_T^{jet} .

3.4 Conclusions

In general, NLO QCD provides a good description of the data. Concerning results on ep -collisions, which are the most interesting ones for this thesis, beauty quark production has been measured using the ZEUS and H1 detector. At the ZEUS experiment, beauty photoproduction has been measured using semileptonic decay into muons. The beauty quarks were identified using the p_T^{rel} method. The measured cross section and the theoretical predictions are found to be in good agreement within errors. The result is still dominated by the statistical errors. At H1, similar results are provided based on an inclusive analysis using the impact parameter method. A light tendency of the data to lie above the NLO prediction is seen here.

This thesis deals with the first inclusive analysis performed so far at ZEUS using a secondary vertex tagging method. The higher statistics of the sample used for the analysis leads to an important reduction of the statistical errors compared to the published ZEUS results. Moreover, given the completely different methods used for the identification of the b -quark, the main systematics uncertainties are for both analyses of different origin, making of the analysis presented here an excellent cross check to the previous result.

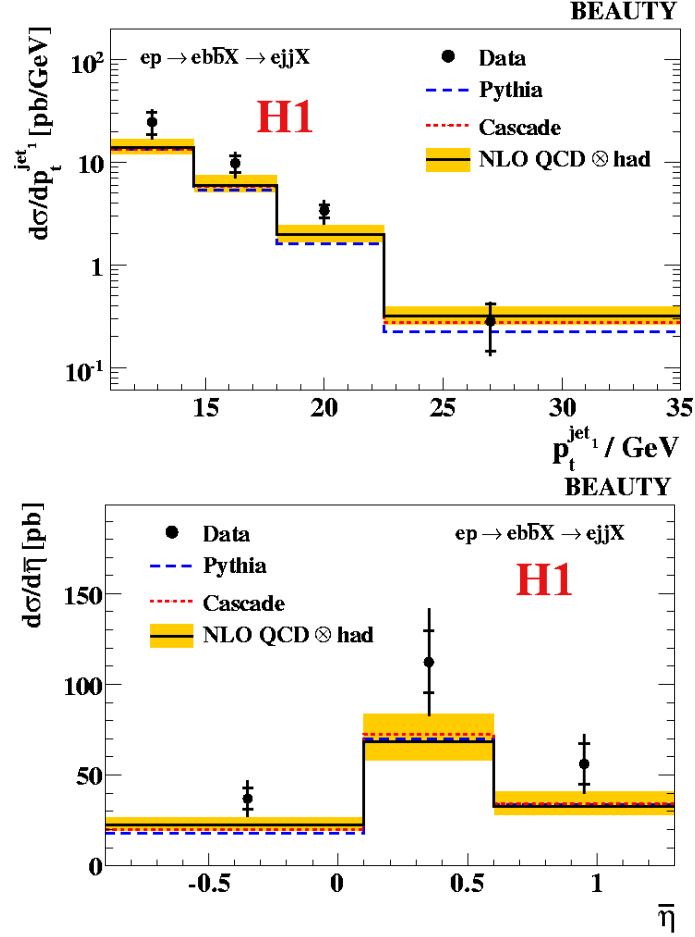


Figure 3.8: Differential beauty photoproduction cross section (top) $d\sigma/dp_T^{\text{jet}}$ and (bottom) $d\sigma/d\eta^{\text{jet}}$ for the process $ep \rightarrow eb\bar{b}X \rightarrow ejjX$. The inner error bars show the statistical errors while the outer error bars indicate the statistical and systematic uncertainties added in quadrature. The measurements are compared to NLO QCD predictions corrected for hadronization effects. The shaded band shows the estimated uncertainty on the theoretical prediction.

Chapter 4

The ZEUS detector at HERA

In this chapter the HERA accelerator and the ZEUS detector are introduced briefly, with particular emphasis on the parts of the detector used in this analysis. A detailed description of the detector can be found in [92].

4.1 The HERA collider

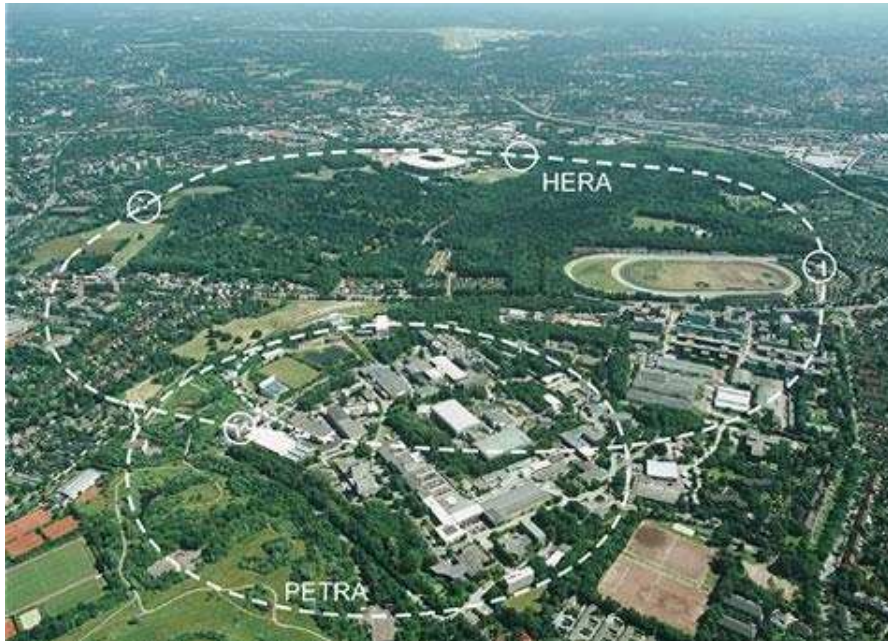


Figure 4.1: Airplane view of the DESY site in Hamburg. The location of the HERA collider and the pre-accelerator PETRA are marked by dashed lines.

The HERA (“**H**adron **E**lektron **R**ing **A**nlage”) is the first and so far the only lepton-proton collider existing in the world [49]. It has been built at the *Deutsches Elektronen Synchrotron* laboratory (DESY) in Hamburg, Germany (Fig. 4.1). The HERA ring has a circumference of 6.3 km and it is located 15-30 m under ground level. Electrons at an energy of 27.5 GeV are brought into collision with protons at 920 GeV (the energy of the proton beam was changed at beginning of 1998 from 820 to 920 GeV) yielding a center of mass energy of 318 GeV.

The two particle beams, accelerated and stored in two separate rings, collide at zero crossing angle at two interaction points, one in the North Hall where the H1 experiment is located, the other in the South Hall where the ZEUS experiment is placed. In addition, there are two fixed-target experiments which make use only of one beam. The HERA-B experiment, located in the West Hall, was used until 2003 to collide the proton beam halo with wire targets and was designed to study CP violation in the $B^0\bar{B}^0$ system. In the East Hall, the HERMES experiment studies the spin structure of the nucleon using the collisions of longitudinally polarized leptons on an internally polarized gas target (H , 2D or 3He).

Figure 4.2 shows a layout of the HERA facility and its pre-accelerator systems. Leptons are first accelerated up to 250 MeV (electrons) and 450 MeV (positrons) in the linear accelerators LINAC I and LINAC II respectively. They are then transferred to the DESY II synchrotron, accelerated up to 7.5 GeV and injected into PETRA II where they are ramped up to 14 GeV and injected into HERA. Once in HERA the leptons are accelerated to the nominal lepton beam energy of 27.5 GeV. The proton acceleration chain starts with negative hydrogen ions (H^-) accelerated to 50 MeV in the proton LINAC. The electrons are stripped off the H^- ions to obtain protons, which are transferred to DESY III and accelerated up to 7.5 GeV and then transferred to PETRA where they are accelerated up to 40 GeV. Finally they are injected into the HERA storage ring, where they reach the nominal beam energy of 920 GeV.

The beams consist of up to 210 bunches. The separation between these bunches is 96 ns and the bunch crossing rate at HERA is 10.4 MHz. Not all bunches are filled. Unpaired or non-colliding bunches are used to study background conditions. When either a lepton or proton bunch is empty, the beam related background can be studied. This background originates from the interactions of the lepton or the proton beam with the residual gas in the beam pipe. When both bunches are empty, non-beam-related background can be estimated and used for calibration purposes or cosmic rays measurements. A summary of HERA design parameters ([49]-[89]) during

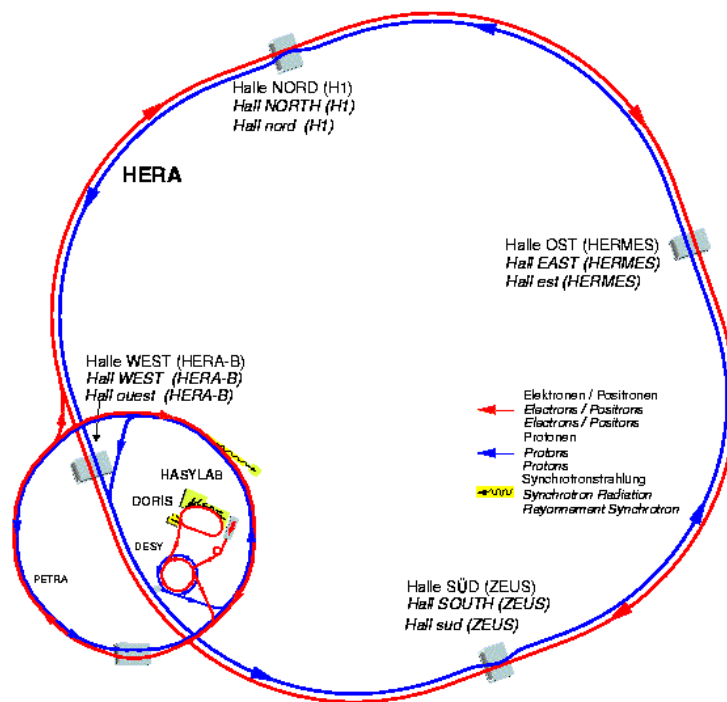


Figure 4.2: Layout of HERA and pre-accelerators.

the different running periods can be found in table¹ 4.1.

Running period	1993 – 1997		1998 – 2000		2003 – 2004	
Luminosity \sqrt{s}	$1.6 \cdot 10^{31} \text{ cm}^{-1} \text{ s}^{-1}$ 300 GeV		$1.6 \cdot 10^{31} \text{ cm}^{-1} \text{ s}^{-1}$ 318 GeV		$7 \cdot 10^{31} \text{ cm}^{-1} \text{ s}^{-1}$ 318 GeV	
Energy	Lepton	Proton	Lepton	Proton	Lepton	Proton
27.5 GeV	27.5 GeV	820 GeV	27.5 GeV	920 GeV	27.5 GeV	920 GeV
Max.nr bunches	210	210	210	210	210	210
Beam current	58 mA	163 mA	58 mA	163 mA	58 mA	140 mA
Part. per bunch	$3.65 \cdot 10^{10}$	10^{11}	$3.65 \cdot 10^{10}$	10^{11}	$4.18 \cdot 10^{10}$	10^{11}
Beam width σ_x	0.286 mm	0.280 mm	0.286 mm	0.280 mm	0.118 mm	0.118 mm
Beam height σ_y	0.060 mm	0.058 mm	0.060 mm	0.058 mm	0.032 mm	0.032 mm

Running period		2005	
Luminosity		$7.44 \cdot 10^{31} \text{ cm}^{-1} \text{ s}^{-1}$	
Center of mass energy		318 GeV	
		Lepton	Proton
Energy		27.5 GeV	920 GeV
Max.nr of bunches²		174	174
Beam current		58 mA	140 mA
Beam width σ_x		< 0.128 mm	0.107 mm
Beam height σ_y		0.027 mm	0.029 mm

Table 4.1: HERA design parameters [49]-[89]

Built between 1984 and 1990, the HERA collider started operation in 1992 in the initial configuration with 820 GeV protons and 27.5 GeV positrons, yielding an ep center-of-mass energy of 300 GeV. In 1998, the maximal proton energy was increased to 920 GeV. Electrons were used during the years 1998, part of 1999, 2005 and part of 2006, positrons in the remaining periods. The delivered luminosity in these configurations is shown in figure 4.3 .

4.1.1 HERA luminosity upgraded

Starting in 2000/2001 the HERA collider was upgraded to achieve a higher specific luminosity. The aim of the upgrade was to increase the design luminosity from $1.5 \cdot 10^{31} \text{ cm}^{-2} \text{ s}^{-1}$ up to $7.4 \cdot 10^{31} \text{ cm}^{-2} \text{ s}^{-1}$. Also the experiments were upgraded with new detector components to provide, together with the increased data sample, higher precision measurements. The luminosity of HERA is defined as:

$$\mathcal{L} = \frac{N_p I_e \gamma_p}{4\pi e \epsilon_N \sqrt{\beta_{xp}^* \beta_{yp}^*}} \quad (4.1)$$

where N_p is the number of protons per bunch, I_e is the total beam current of the lepton beam, γ_p is the Lorentz factor for the protons, e is the elementary charge and $\beta_{xp,yp}^*$ are the horizontal and vertical beta functions at the

¹Because of technical reasons, the used max.nr of bunches was 180

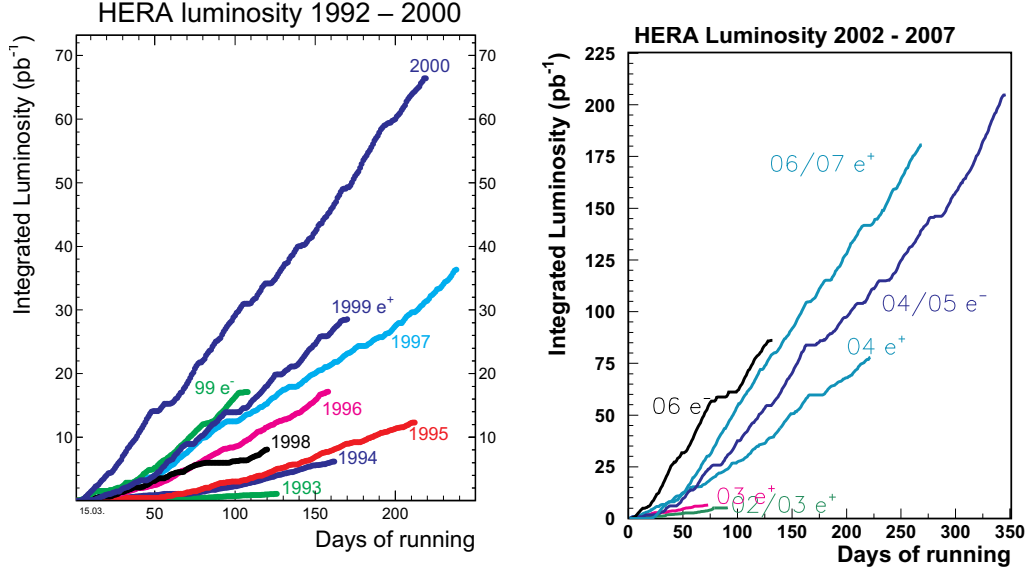


Figure 4.3: Integrated luminosity delivered by HERA for the running periods 1993-2000 (left) and 2002-2007 (right)

interaction point.

Due to the fact that the total lepton beam current I_e is limited by the available radio-frequency(RF) power and the maximum proton density N_p is limited by the space charge effects in DESY III, the HERA luminosity could be increased only by removing the limit on the proton beam β -functions. The beam cross sections at the interaction point (IP) were decreased by stronger focusing of both the electron and the proton beam. This required new superconducting magnets close to the IP's as well as an absorption system for the increased synchrotron radiation emitted inside the experiment region. The HERA upgrade is described in detail in the HERA reports [88] and [64].

4.2 The ZEUS detector

The ZEUS³ detector [92] is a multipurpose detector to study lepton proton scattering at HERA. It is a quasi-hermetic detector as it covers most of the 4π solid angle with the exception of small regions around the beam pipe. The higher energy of the protons relative to the electrons leads to a boost of the center-of-mass system relative to the laboratory system. Consequently,

³ZEΥΣ : Ζήτησις καθ' Ευρετής Υποκειμένης Συμμετρίας (Search to Elucidate Underlying Symmetry)

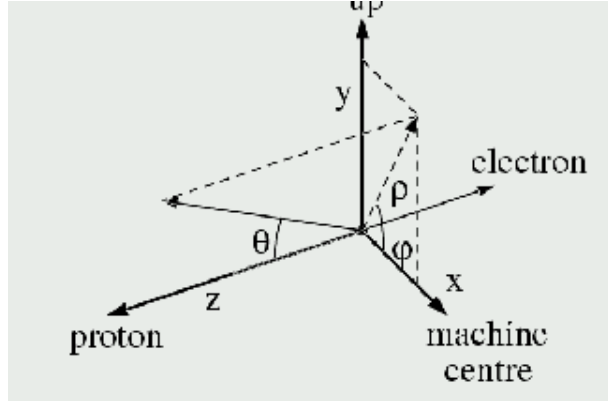


Figure 4.4: The ZEUS coordinate system

the ZEUS detector presents an asymmetric topology with additional forward tracking detectors in the proton direction.

The ZEUS coordinate system, depicted in Fig. 4.4, is a right-handed orthogonal system with the origin at the nominal interaction point, the z-axis pointing in the proton direction (also referred as forward direction), the x-axis pointing toward the center of HERA and the y-axis pointing upward. The polar angle θ and the azimuthal angle ϕ are measured relative to the z and x axes respectively. Often the pseudorapidity variable, $\eta = -\ln \tan(\theta/2)$, is used instead of θ .

The main detector is approximately 20 m long, 12 m wide and 11 m high and has a weight of 3600 tons. An overview of the detector in the z-y and the x-y planes can be seen in Fig. 4.5 and 4.6 respectively. A brief outline of the main components is given below followed by a more detailed description of the sub-detectors relevant for the analysis presented in this thesis.

Directly around the interaction region lie the charged particles tracking detectors. The innermost detector was the Vertex Detector (VXD) which was removed during the 1995-1996 shutdown and, in 2001, a silicon micro-vertex detector (MVD) was installed in its place. The main tracking device is the CTD (Central Tracking Detector) which is a cylindrical drift chamber surrounding the beam pipe at the interaction point. The CTD is supplemented in the forward region with three sets of planar drift chambers (FTD) with interleaved Transition Radiation Detectors (TRD), and in the rear direction with one planar drift chamber consisting of three layers (RTD). In 2001, the TRD was replaced by an additional forward tracking detector called Straw Tube Tracker (STT). The tracking system is enclosed by a superconducting solenoid providing a magnetic field of 1.43 T for the determination of charge

and momentum of charged particles.

A high-resolution compensating uranium-scintillator calorimeter (CAL) surrounds the tracking detectors and the magnet. It is the main device for energy measurements and is divided into the forward (FCAL), rear (RCAL) and barrel (BCAL) sections. Attached to the front face of the RCAL, is the small angle rear tracking detector (SRTD) which is used to improve the position measurement of charged particles within its acceptance and particularly for a precise determination of the scattered electron impact position. The CAL is surrounded by the iron yoke which provides a return path for the solenoid magnetic field flux and is equipped with proportional tubes serving as calorimeter (BAC) for the particles which have not been absorbed by the CAL. The BAC is also used as muon detection system. Dedicated muon identification detectors are located inside (FMUI, BMUI and RMUI) and outside (FMUO, BMUO, RMUO) the iron yoke.

The VETO wall is located at $z = -7.5$ m from the interaction point in the rear direction. It consists of an iron wall supporting scintillator hodoscopes and is used to reject background from beam-gas interactions. The LUMI detector is made of two electromagnetic calorimeters placed at $z = -34$ m and $z = -107$ m to measure the electrons and photons from bremsstrahlung events for the determination of the luminosity.

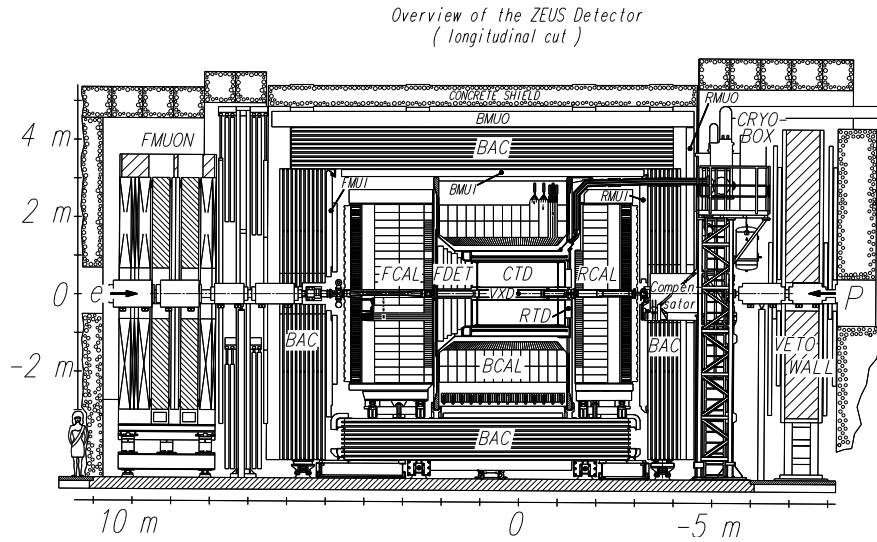


Figure 4.5: Cross section of the ZEUS detector along the beam axis

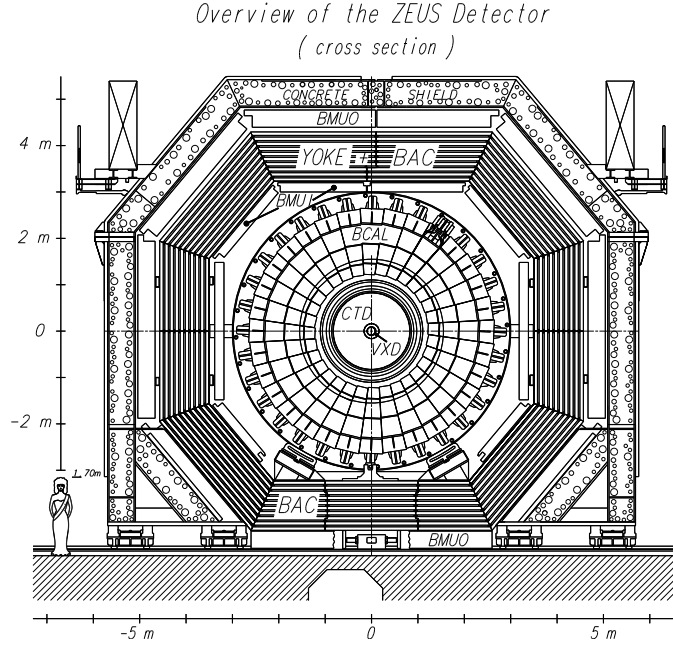


Figure 4.6: Cross section of the ZEUS detector perpendicular to the beam axis.

4.2.1 The Central Tracking Detector

The Central Tracking Detector [18] is a cylindrical drift chamber which measures the direction and momentum of charged particles and the energy loss dE/dx , which provides information for particle identification. The inner radius of the chamber is 18.2 cm and the outer radius is 79.4 cm. It covers the longitudinal interval from $z = -100$ cm to $z = 103$ cm resulting in a polar angle coverage of $15^\circ < \theta < 164^\circ$. The CTD is filled with a mixture of argon (Ar), carbon dioxide (CO_2) and ethane (C_2H_6) in the proportion 90:8:2.

When a charged particle crosses the CTD, it produces ionization of the gas in the chamber. The electrons from the ionization drift towards the positive sense wires, whereas the positively charged ions will move to the negative field wires. The drift velocity of the electrons is approximately constant and equal to $50 \mu m/ns$. An avalanche effect occurs close to the wire yielding an amplification factor for the electrons of about $\sim 10^4$ so that a measurable pulse is induced on the sense wires which is measured via electronic read-out.

An octant of the CTD is shown in figure 4.7. The CTD wires are arranged in nine superlayers (SL). Each SL consists of cells made up of 8 sense wires with its associated field wires (see Fig. 4.8). The sense wires are $30 \mu m$ thick while the field wires have different diameters. The CTD is designed to

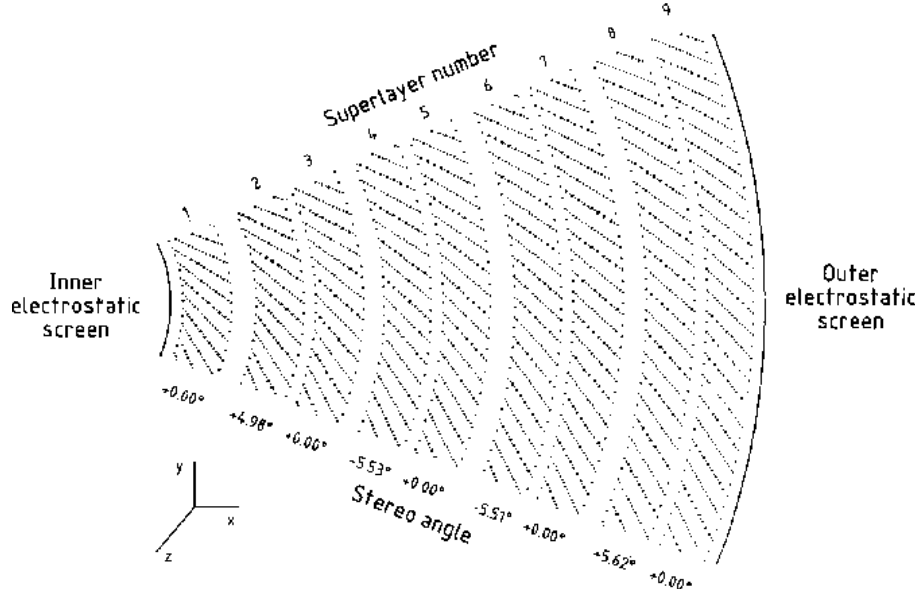


Figure 4.7: A x-y cross section of one octant of the CTD. The sense wires are indicated with dots.

operate in a magnetic field to improve the momentum measurement of the charged particles. The Lorentz angle of the magnetic field is 45° with respect to the radial axis. To compensate this, the sense wires planes are tilted with respect to the radius.

The SLs are numbered from 1 (innermost) to 9 (outermost). Odd numbered SLs have wires parallel to the beam axis (axial superlayers), whereas the wires of the even numbered SLs are inclined by a stereo angle of $\pm 5^\circ$ (stereo layers). This allows an accurate determination of the z-position of the hit. The achieved resolution is $\sim 100 - 120 \mu m$ in the $r - \phi$ plane and $\sim 1.4 mm$ in the z coordinate.

Additionally, the layers 1, 3 and 5 are instrumented with a z-by-timing system which determines the z-coordinates of a hit by measuring the difference in arrival time of the pulse at both ends of the CTD. The resolution achieved on the z-coordinate with this system of $\sim 4 cm$, is worse than the one obtained using full axial and stereo wire information but it is faster and, therefore, used predominantly for trigger purposes.

The resolution on the transverse momentum p_T [28] of tracks fitted to the interaction vertex and passing at least three CTD layers with $p_T > 150 MeV$ is given by:

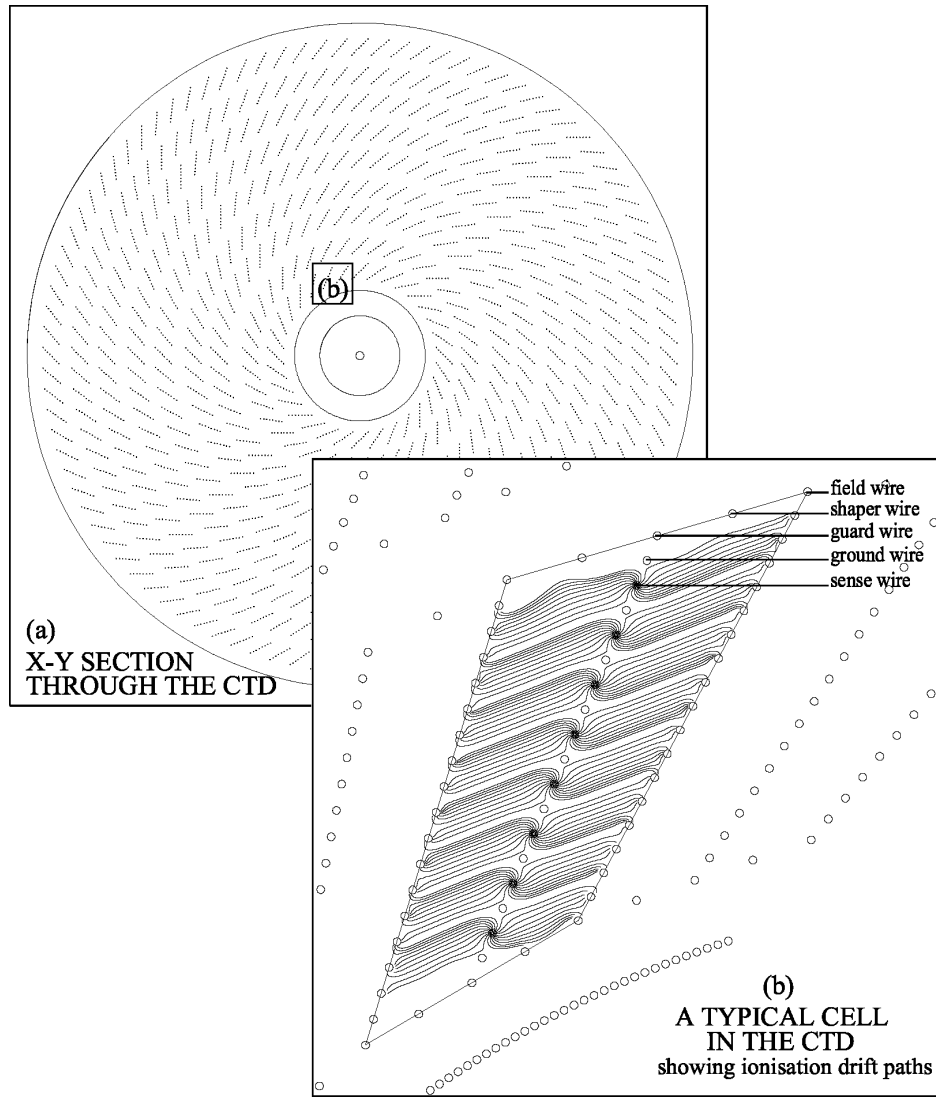


Figure 4.8: (a) $x - y$ view through CTD and (b) a zoomed single cell with field and signal wires

$$\frac{\sigma(p_T)}{p_T} = 0.0058 \cdot p_T \oplus 0.0065 \oplus \frac{0.0014}{p_T}, \quad (4.2)$$

where p_T is expressed in GeV and the symbol \oplus indicates the quadratic sum. The first term is due to the hit position resolution, while the second and the third are due to multiple scattering before and inside the CTD respectively.

Since the installation of the MVD in 2001, the resolution has changed and the influence of the new detector components on the resolution of the tracking system is partially still under investigation. The influence from multiple scattering is larger due to the extra material of the MVD whereas the hit resolution is improved thanks to the inclusion of the MVD hits. The transverse momentum resolution, as determined at the end of 2005 [17], including the MVD in the global track reconstruction indicate is:

$$\frac{\sigma(p_T)}{p_T} = 0.0026 \cdot p_T \oplus 0.0104 \oplus \frac{0.0019}{p_T}, \quad (4.3)$$

Comparing it to the corresponding resolution without MVD (4.2) it is seen that for low momentum tracks the resolution is similar but for high momentum tracks an improvement of a factor two is achieved. There has been considerable improvement since this last value. Though there is not a parametrization yet, the latest studies on momentum resolution can be found in [79].

4.2.2 The Micro Vertex Detector

The silicon-strip micro vertex detector (MVD), installed in 2001 during the HERA upgrade shutdown, was designed to improve the overall precision of the tracking system and allow the identification of events containing secondary vertices from decays of long-lived particles ($c\tau > 100\mu m$) [13]. The physics requirements and the limited available space given by the size of the beampipe and the inner volume of the CTD, lead to the following detector design characteristics:

- angular coverage of $10^\circ - 160^\circ$;
- three spatial measurements, in two projections each, per track;
- $25\mu m$ intrinsic hit resolution;
- two-track separation of $200\mu m$;
- hit efficiency better than 95%.

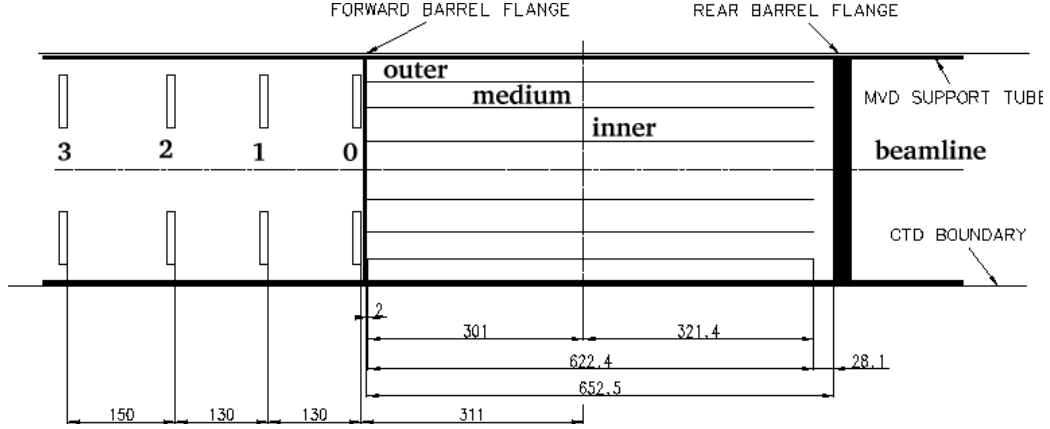


Figure 4.9: Longitudinal MVD section: There are three layers of silicon sensors (*ladders*) parallel to the beam line in the barrel region (BMVD) and 4 layers perpendicular to the beam line (*wheels*) in the forward region (FMVD)

For a good matching of the silicon planes with the existing tracking detectors (CTD, FDET and RTD), the MVD is split into two sections: barrel (BMVD) and forward (FMVD) (see Fig. 4.9).

The BMVD is 64 cm long and covers a polar angle between $30^\circ - 150^\circ$. It consists of three layers of silicon sensors arranged in concentric planes around the interaction point. The inner layer, placed at a variable radius between 3 and 5 cm from the CTD axis, is not complete due to the elliptical shaped beam pipe which is not centered around the interaction point (see figure 4.10). The second and third layers are placed at $r \sim 8.6$ cm and $r \sim 12.3$ cm.

The BMVD is equipped with 600 single-sided strip silicon sensors of approximately 64×64 mm² and 320 μ m thick with p^+ strips implanted into n-type bulk. Each one is covered with 3082 strips with a pitch of 20 μ m. Two sensors are glued together next to each other with perpendicular strips on the same side. When placed in the barrel MVD, the sensors parallel to the beam pipe give the $r - \phi$ coordinate while the perpendicular ones give the z information. The two sensors are read out together bringing the strips of the two planes with an interconnection circuit into contact. A fan-out circuit connects the readout strips to the front-end electronics, placed on the MVD hybrids. Every sixth strip is read out resulting in 512 readout strips at a readout pitch of 120 μ m. Two symmetric assemblies are produced, called right and left *half-module*. The two half-modules are finally placed on top of each other forming a readout *full module*. Figure 4.11 shows the schematic view of a half module as well as how they are combined to form a module. Five full modules are fixed side by side, with 1 mm gap in between, on a 30

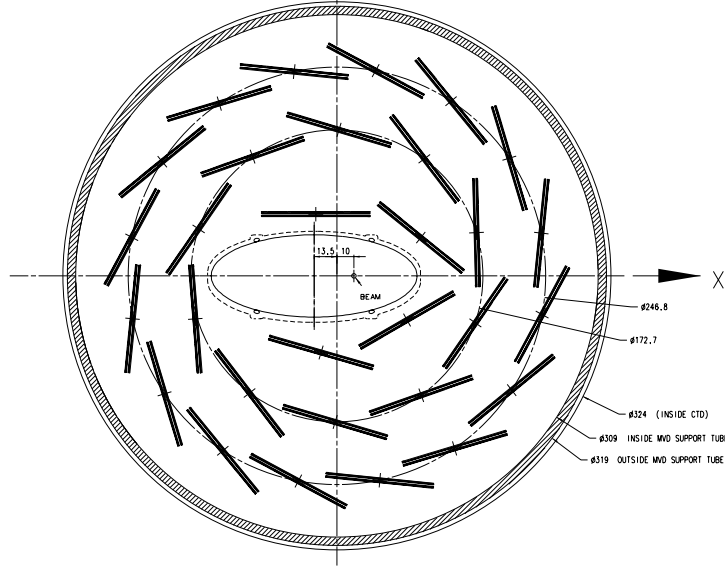


Figure 4.10: BMVD section: silicon detectors layers without support structures are shown. The layers are positioned around the beam pipe.

carbon fibre structure, and form a *ladder*.

The FMVD consists of four planes (*wheels*) perpendicular to the beam axis (Fig. 4.12). The four wheels are positioned at $z=32, 45, 47$ and 75 cm. Each wheel is made of two back to back layers of 14 silicon sensors of the same type as in the barrel section but with a trapezoidal shape. The FMVD extends the polar coverage up to 7° providing additional forward tracking capabilities. In each wheel the two layers are parallel but the strips are tilted by $\sim 13^\circ$ ($180^\circ/4$) in opposite directions to provide two coordinates for a particle traversing the wheel. The number of readout strips per sensor is 480.

The MVD silicon sensors are read out using the analog chips HELIX 3.0 [38] which are mounted on ceramic structures called the hybrids (see figure 4.11). One HELIX chip has 128 input channels. The front-end of the chip consists of a charge sensitive preamplifier, a shaper and a buffer. Signals from the silicon detectors are first amplified and shaped in a pulse with a peak height proportional to the measured charged. Then, a discriminator (GFLT, see 4.2.5) determines for each bunch crossing if the signal from the detector is above or below a given threshold (trigger decision). The HELIX outgoing signal is transferred with an analog connection to the ADC boards, which perform a first cluster reconstruction. The signal is then transferred to the MVD second level trigger processor and to the ZEUS event builder.

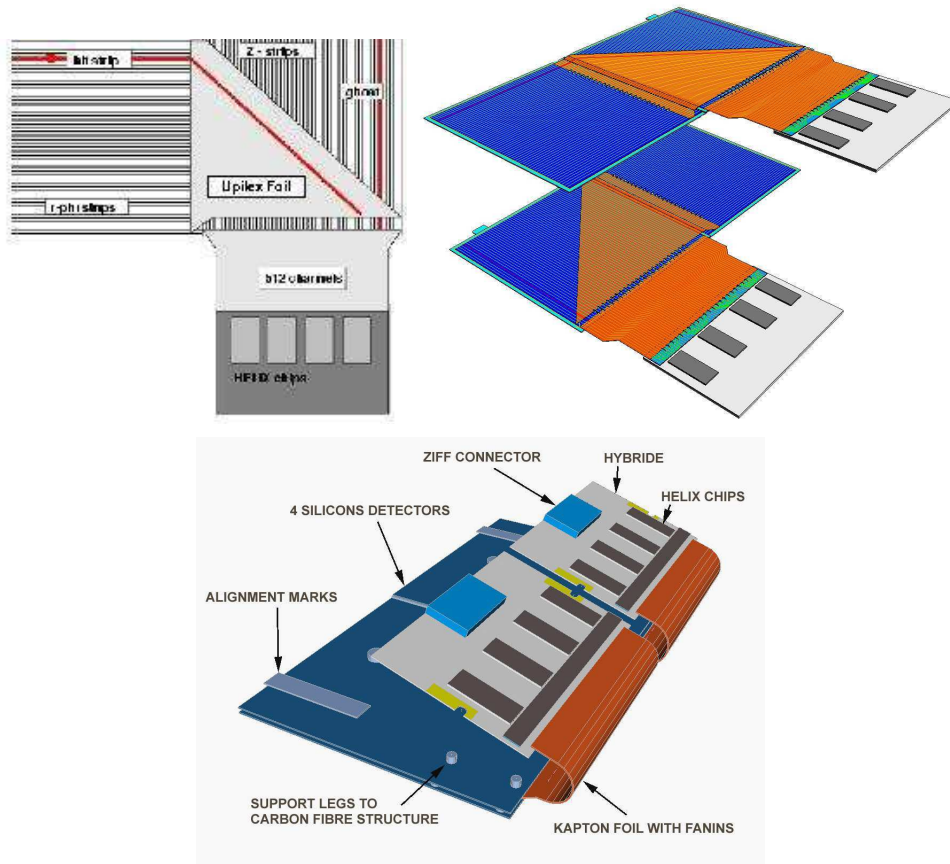


Figure 4.11: Schematic view of a half-module (top left), assembly of two half-modules into a module (top right) and a ladder mounted on the support (bottom)

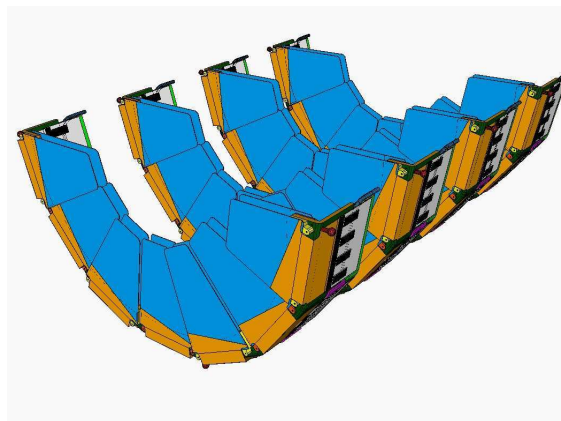


Figure 4.12: FMVD 4 wheels

Radiation Monitor

A special feature of the MVD is that the readout chips are located in the sensitive region. Hence, they are exposed to radiation. In order to keep the radiation dose under control, a radiation monitoring and automatic beam-dump system (Radmon) has been installed. The main radiation sources in the MVD area are the synchrotron radiation and the radiation caused by beam losses. The maximum tolerable dose of ionizing radiation for the detector and its readout electronics is 3 kGy, which is approximately five times higher than the average dose rate in the HERA-I data taking periods. The dose rate measurement is based on the readout of the radiation induced signal currents of silicon PIN diodes. A detail description of this system can be found in [25].

4.2.3 The Uranium-Scintillator Calorimeter

The ZEUS calorimeter (UCAL) [15] is a high-resolution compensating sampling calorimeter. It completely surrounds the tracking devices and covers 99.7% of the 4π solid angle. It consists of alternated layers of absorbing and scintillating material. The absorber layers are made of 3.3 mm thick depleted uranium plates (98.1% U^{238} , 1.7% Nb, 0.2% U^{235}) wrapped in stainless steel foils. The active layers are made of 2.6 mm thick scintillator (SCSN38) plates.

Uranium was chosen as absorber material since it provides a high yield of spallation neutrons in hadronic interactions which can be detected with the plastic scintillators giving an increased signal. This compensates the signal loss of hadronic interactions due to losses in binding energy, nuclear fission and decay products escaping detection. These losses are not present in interactions of electrons or photons since they lose their energy predominantly by interactions with atomic electrons. Compensation can be achieved by the right tuning of the relative thickness of the active and the absorber material in order to have the same response for an electron and a hadron of the same energy passing through the detector:

$$e/h = 1.00 \pm 0.05. \quad (4.4)$$

The electromagnetic and hadronic energy resolutions measured under test-beam conditions are:

$$\frac{\sigma_e}{E} = \frac{18\%}{\sqrt{E}} \oplus 2\% \text{ and } \frac{\sigma_{had}}{E} = \frac{35\%}{\sqrt{E}} \oplus 1\% \quad (4.5)$$

where E is the particle energy measured in GeV.

The UCAL is divided into three parts: the forward (FCAL), barrel (BCAL) and the rear (RCAL) calorimeter (Fig. 4.13). The thickness of each part is different due to the fact that the average final state particles in a lepton proton interaction at HERA are boosted into the forward (proton) direction. The FCAL is $\sim 7 \lambda$, the BCAL is $\sim 5 \lambda$ and the RCAL $\sim 4 \lambda$, where λ is the interaction length⁴. Each part of the calorimeter is divided into *modules* which are groups of towers ($20 \times 20 \text{ cm}^2$) containing stacks of absorber and scintillator material. Each module is segmented longitudinally into one electromagnetic (EMC) and two (one in RCAL) hadronic (HAC) sections. The EMC sections are further divided into four *cells* (two in RCAL), the smallest subdivision of the calorimeter. Figure 4.14 shows a layout of a FCAL module and the tower structures in FCAL and RCAL.

The BCAL has 32 wedge-shaped modules corresponding to its cylindrical geometry. The EMC section has a depth of $21 X^0$, where X^0 is the radiation length and each of the HAC sections is 2λ deep. The FCAL and RCAL consist of 23 rectangular modules. The resulting depth is $26 X^0$ for the electromagnetic section and 3.1λ for each hadronic section.

The scintillator light of each cell is collected by light guides and wavelength shifters on both sides of the module and brought to one of the 11386 photomultiplier tubes (PMT) where it is converted into an electrical signal. The comparison of the two signals can be used to determine the horizontal position of the shower in the cell. The sum of the signals of both PMTs is independent of the impact parameter of the particle and used as time and energy measurement with a resolution of the order of 1 ns (for energy depositions larger than 1 GeV). This excellent time resolution is used to reject non *ep*-background (e.g. particles from beam halo) already at the trigger level (see section 4.2.5 for description of the trigger chain).

The CAL is calibrated on a daily basis using the stable signal from the radioactive decay of the depleted Uranium. The PMTs and the readout electronics are additionally calibrated using charge, LED, laser and test pulses.

⁴The nuclear interaction length λ is defined as the average distance that strongly interacting particles travel in a material such that $\lambda = A/(N \cdot \sigma \cdot \varrho)$, where N is Avogadro's number, A is the atomic mass of the medium, ϱ is the density and σ the absorption cross section.

The radiation length X^0 is defined as the characteristic amount of matter traversed by high-energy electrons or high-energy photons. It is both (a) the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung, and (b) $7/9$ of the mean free path for pair production by a high-energy photon

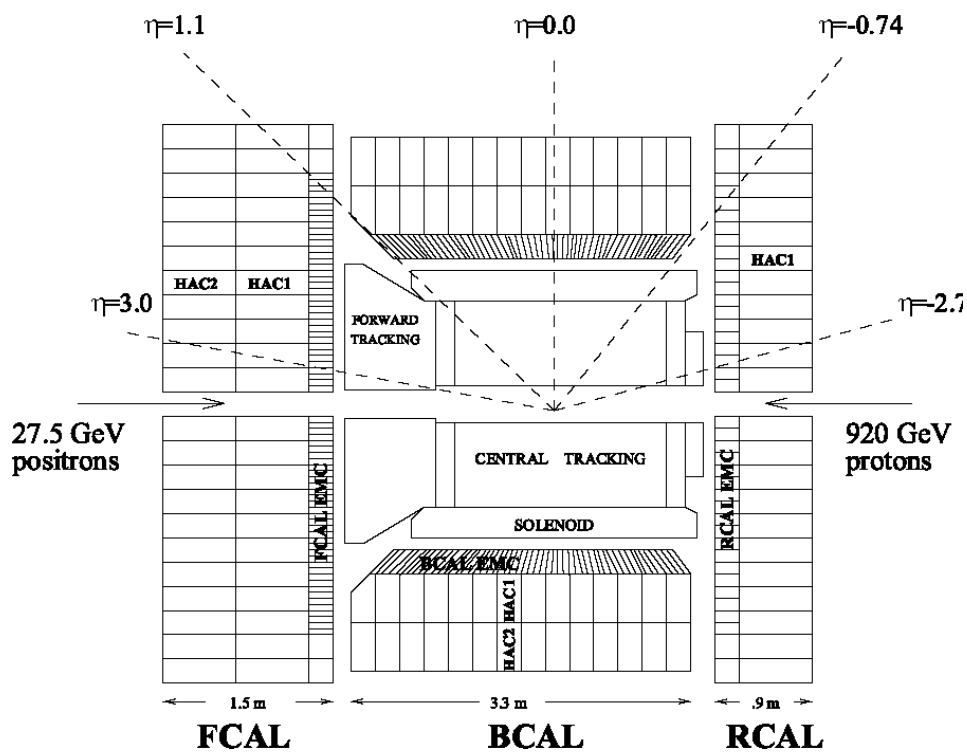


Figure 4.13: Schematic view of the UCAL along the beam axis.

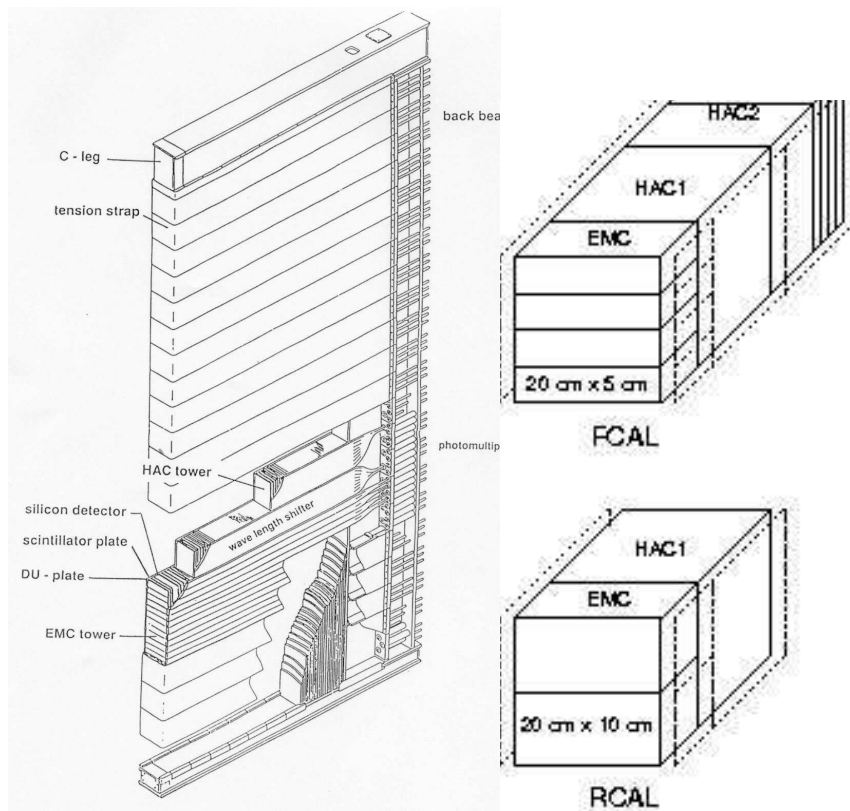


Figure 4.14: View of FCAL module (left). Tower structure in FCAL and RCAL (right)

The achieved accuracy is better than 2%.

4.2.4 The Luminosity Measurement

The luminosity at ZEUS [6] is determined by measuring the rate of photons which emerge from the Bethe-Heitler process $ep \rightarrow ep\gamma$, where the lepton and the photon are scattered at very small angles. The cross section of this process is large ($\sigma_{BH} \geq 20 \text{ mb}$) providing enough statistics and can be calculated as a function of the photon energy with an accuracy of 0.5% [48]. This can be used for a precise determination of the luminosity using the relation:

$$\mathcal{L} = \frac{N_\gamma}{\sigma_{BH}} \quad (4.6)$$

where N_γ is the corresponding observed number of bremsstrahlung photons.

The luminosity monitor system consists of a photon and a lepton lead-scintillator sampling calorimeter, located along the beam line at $z = -107 \text{ m}$ and $z = -35 \text{ m}$ respectively (Fig. 4.15). Photons with a scattering angle $\theta_\gamma < 0.5 \text{ mrad}$ exit the beam pipe and strike the photon calorimeter. Leptons with an energy between $0.2E_e$ and $0.8E_e$ are deflected away from the beam line by magnets and hit the lepton calorimeter. A carbon-lead filter shields the detector from synchrotron radiation. The resulting calorimeter resolution is $\sigma(E)/E = 23\%/\sqrt{E}$. The bremsstrahlung event rate is determined by counting the number of photons above a fixed energy threshold.

The main contribution to the background is given by the bremsstrahlung of the leptons on the residual gas in the beam pipe. This can be estimated using *pilot* bunches (lepton bunches with no corresponding proton bunch), evaluating for these the rate of bremsstrahlung events. The achieved precision on the luminosity measurement for 2004 data is of the order of 4.2 – 4.7% [52].

4.2.5 The ZEUS Trigger System

The aim of the ZEUS trigger system [92] is to select the events of interest from the ep interactions with a high, stable and well understood efficiency and, at the same time, reject background events and reduce the flow of information resulting from the high bunch-crossing rate in order to be able to write events on tape with the highest possible efficiency. HERA beam bunches cross every 96 ns which corresponds to a bunch crossing rate of 10 MHz. The rate of background, coming mainly from beam-gas interactions

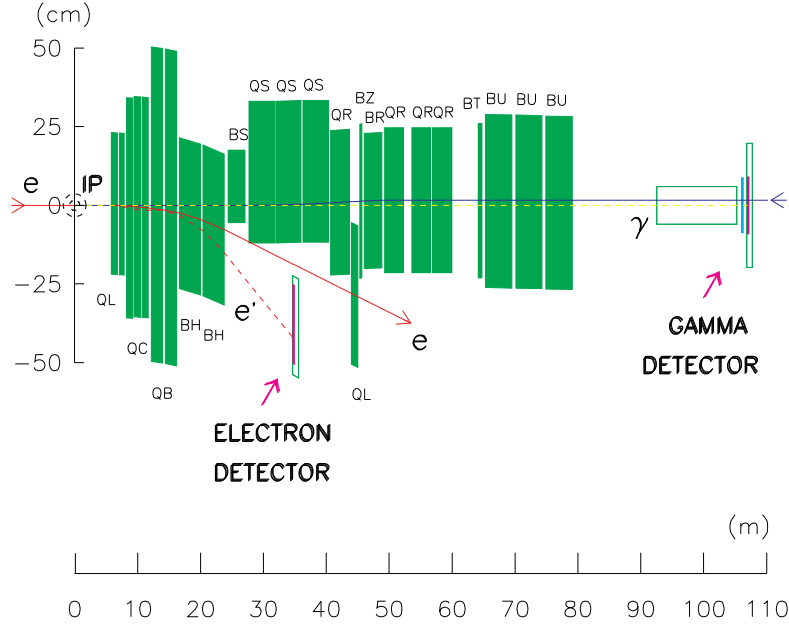


Figure 4.15: Schematic view of the ZEUS luminosity monitor system. The photon detector is situated at $z=-107$ m and the lepton detector at $z=-35$ m

and cosmic rays, is about 10 kHz, which exceeds the rate of physics events⁵ by several orders of magnitude. The total data size per event is 150 kB and the writing speed is limited to ~ 1.5 MB/s making a significant reduction of the data rate and size necessary.

The ZEUS trigger system is organized in three levels, as described in the following sections. A schematic diagram of the trigger and the data acquisition system is shown in figure 4.16.

First Level Trigger (FLT)

Each detector component is equipped with its own FLT, implemented in hard-wired logic circuits and designed to analyze one event per bunch crossing. The data of every bunch crossing is stored in pipelines which store up to 46 bunch crossings. Each FLT provides a fast trigger decision based on properties like energy sums, thresholds or timing information. After 26 bunch crossing times ($2.5 \mu s$) all the components send their information to the global first level trigger (GFLT) which makes a final decision within 1.9

⁵The rate of ep events ranges from less than 0.1 Hz for NC DIS events with $Q^2 > 100 \text{ GeV}^2$ to 250 Hz for soft photoproduction for an instantaneous luminosity of $2 \cdot 10^{31} \text{ cm}^{-1} \text{ s}^{-1}$

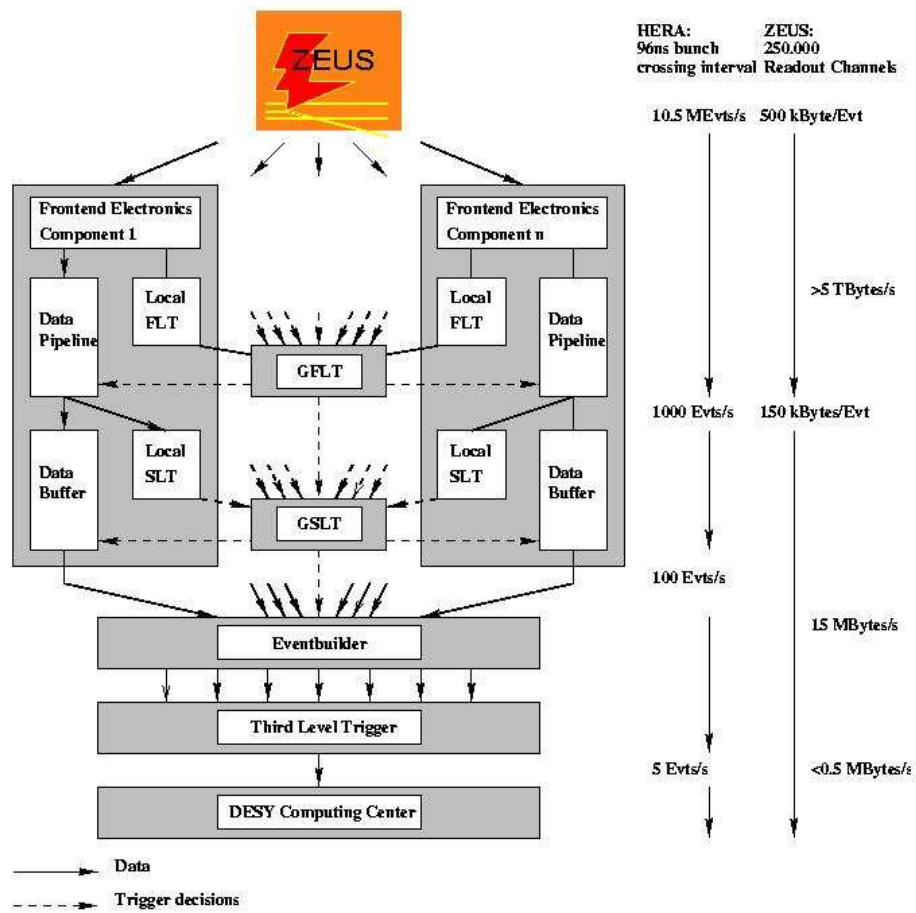


Figure 4.16: The ZEUS trigger and the data acquisition system.

μs (corresponding to the remaining 20 bunch crossing times). The overall time taken by the FLT is hence $\sim 4.4 \mu s$ which corresponds to 46 HERA bunch crossings stored in the pipelines. If the event is accepted, the information is sent to the second level trigger. The typical GFLT output rate is below 1 kHz.

Second Level Trigger (SLT)

The SLT is software based and uses a network of transputers [55] to reduce the event rate from ~ 1 kHz (output rate from GFLT) to below 100 Hz. The events accepted by the GFLT are analyzed further and event quantities are recalculated to a higher degree of precision. CAL timing information, available at this stage, is used to reject beam gas background. Analogous to the FLT, each detector component has its own SLT processor that sends its response to the global second level trigger (GSLT). It provides a decision after 7 ms. For GSLT accepted events, the data from all components is combined in the event builder (EVB) which writes them out in a standard format (ADAMO database tables [74]) and makes them accessible to the third level trigger (TLT).

Third Level Trigger (TLT)

The TLT is a software trigger running on a computer farm. It consists of a series of algorithms (a version of the offline reconstruction software) which reconstructs the event fully using the information from several components. After this reconstruction, the events are classified into physics categories according to various filters set by the physics working groups. After an event is accepted, the TLT sends the data to the DESY computer center where the events are written to tape for a re-processing with complete calibrations and the full reconstruction software. The final output rate of the TLT is $\sim 5 - 15$ Hz.

Chapter 5

Event Simulation

Within high-energy physics, simulation programs have become essential, being used from the design of the experiment up to the analysis of the data. Through comparison between measurements and theoretical predictions, a better understanding of the detector's response to the physics events is achieved. Theoretical models can be tested as well.

The event simulation at HERA consists essentially of two parts:

- Physics process: The ep scattering process is simulated using an *event generator* program. Within the framework of a theory model, this program provides a complete description of the final state particles produced in the primary interaction. This part was described in detail in Section 2.8.
- Detector simulation: All the detector components and trigger systems are simulated. Then, the final state particles generated from the primary interaction in the previous part are propagated through the simulated detector. Section 5.2 describes how the detector simulation is performed.

Since these simulations are based on Monte Carlo (MC) statistical methods, the whole event simulation is commonly referred to as Monte Carlo simulation and the event generators as MC generators. The MC generator used in this thesis is PYTHIA and was described in the chapter 2 of this thesis. Here the global procedure of an event simulation is explained.

5.1 Physics simulation.

The QCD Monte Carlo event generators are constructed on the basis of the factorization theorem [11] which allows the separation of the cross sections

for scattering processes into parts which are soft, or non-perturbative, and hard, or perturbative. The soft components, such as parton density functions (PDF's) and fragmentation functions (hadronization) must be derived from experimental measurements using phenomenological approaches while the hard components like the partonic hard-scattering cross sections, are calculable in pQCD

The event simulation, outlined in Figure 5.1, is performed in the following stages:

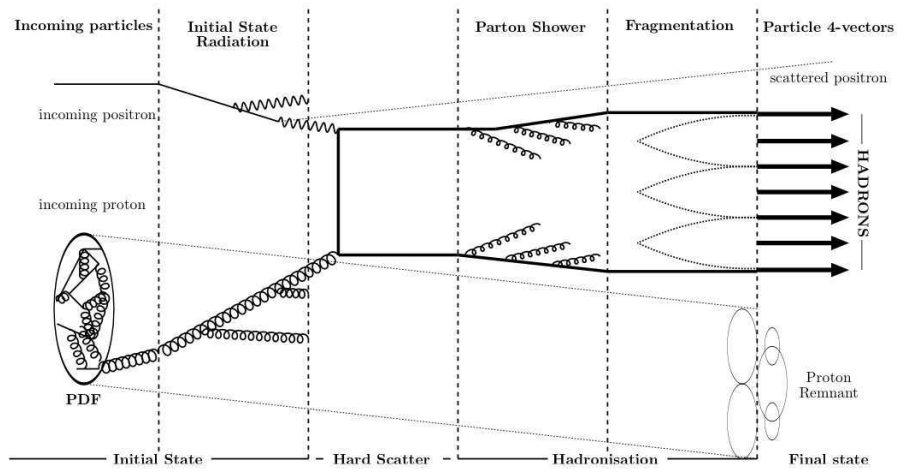


Figure 5.1: General structure of a leading order plus parton shower MC event generator.

- **Hard sub-process:** It is the interaction between a parton from the proton and the exchanged photon, or photon constituent in resolved photon events. This process can be calculated at fixed order perturbative expansion if it involves a hard scale $\mu \gg \Lambda_{QCD}$.
- **Initial and final state radiation:** Partons (also leptons and photons) can emit other partons, in $a \rightarrow bc$ processes, during the initial and final state (before and after the hard scatter respectively). Each daughter b or c of the process can branch again in turn. This parton emission cascade is also known as *parton shower* and contributes to higher order corrections not taken into account by the LO hard subprocesses. These higher order effects are important at high energies where they can have a big influence on the topology of the event (e.g. multiple parton emission can give rise to multijet events). Some of the branchings included in a general MC generator program are: $q \rightarrow qg$, $q \rightarrow q\gamma$,

$g \rightarrow q\bar{q}$, $g \rightarrow gg$ and $l \rightarrow l\gamma$. The kinematics of each branching is given by two variables: Q^2 which is the transverse momentum squared of the branching and z defined as the fraction of parent parton energy carried by one of the daughters, being $(1 - z)$ the fraction kept by the other daughter. The probability $P_{a \rightarrow bc}(z)$ to branch is given by the DGLAP equations [4] in the leading approximation of perturbative QCD.

- **Hadronization/Fragmentation:** It is the process in which, due to color confinement, colorless hadrons are formed out of colored partons (quarks and gluons). It is a non-perturbative phenomenon and is not yet calculable from first principles. The MC event generator PYTHIA uses a phenomenological model called Lund string model, described in detail in section 2.6.2.
- **Beam remnant:** In ep collisions, the initial-state radiation algorithm reconstructs one shower initiator in each beam i.e. one parton from the incoming proton and the exchanged photon (if a photon constituent in resolved events). These shower initiators, carrying only a fraction of the initial beams energies, will start the initial-state parton showers in which the partons that enter in the hard scatter and in the final-state showers are created. The remnant-beam is thus the incoming particles “minus” those shower initiators. The flavour and color of the beam remnant is reconstructed using the hadronization model, given the flavour and color of the initiator parton.

The general-purpose event generator adopted to generate the MC samples used in this thesis is PYTHIA. As already described in section 2.8, the initial input of PYTHIA are the parton distributions of the colliding leptons and protons. In particular, for photoproduction events the Weizsäcker Williams Approximation was used to generate the spectrum of photons radiated from the incoming lepton. The event simulation procedure will then follow the different stages summarized above, with special emphasis on the hadronization processes, based on the string fragmentation or Lund model.

5.2 Detector simulation

The event generator PYTHIA, like all event generators supported in ZEUS, is included in the AMADEUS package [5]. AMADEUS is a software program whose main aim is to convert the output event record of the event generator, such as the four-momenta of the particles produced in the hard scatter and all the relevant kinematic variables, to a suitable format (ADAMO format)

for the following simulation of the ZEUS detector response and the trigger system. A diagram of the ZEUS reconstruction scheme for data and Monte Carlo is shown in Figure 5.2.

Following the reconstruction chain, the data produced by the event generator are the input to MOZART¹, the ZEUS detector simulation program. MOZART is based on the GEANT [68] package which contains a complete description of the detector, including geometrical shapes, positions, type and amount of material of all detector components in detail. GEANT traces the particles through the detector, simulating its response and taking into account physics processes such as energy loss, multiple scattering and particle decays in flight. The output of MOZART is the input for the CZAR² package, which simulates the trigger logic as used in the data taking. The next step, common for both MC-simulated events and ZEUS raw data, consists on the offline reconstruction which is performed by the package ZEPHYR³.

As mentioned above, the ZEUS data is organized into a suitable format using the ADAMO⁴ data management system [74]. This system is based on the Entity Relationship Model for the design of data structures, for their storage in memory and on external media (tape or disk), and for their documentation. Finally, the users can access the data with the EAZE⁵ program or ORANGE⁶, a more inclusive EAZE job, where both MC-simulated events and data are analyzed in an identical way.

5.3 MC samples used in this analysis

In this Section, all the MC samples generated for the simulation of the data used in the first analysis presented in this thesis are described below and summarized in Table 5.1.

5.3.1 Beauty MC samples

In Chapter 2, the different production mechanisms existing for beauty quarks production in ep collisions were described. According to those, the beauty quark samples were generated separately for four sub-processes using the PYTHIA MC generator:

¹**MO**nTe Carlo for **Z**eus **A**nalysis **R**econstruction and **T**riggering

²**C**omplete **Z**gana **A**analysis **R**outine

³**ZE** **PHY**sics **R**econstruction

⁴**A**leph **D**ata **M**odel

⁵**E**ffortless **A**nalysis of **Z**eus **E**vents

⁶**O**verlying **R**outine for **A**analysis **N**tuple **G**eneration

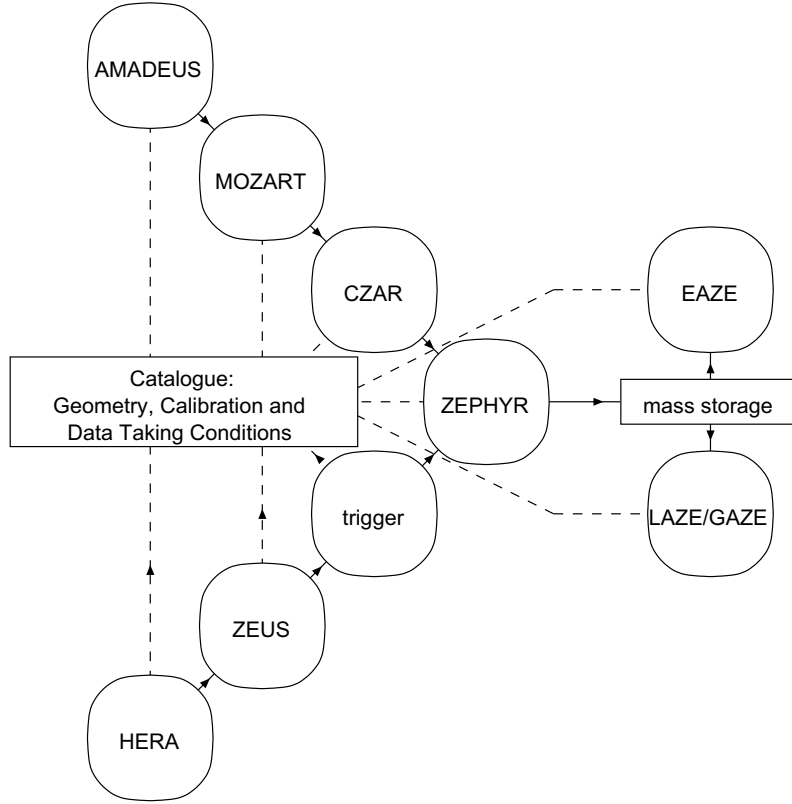


Figure 5.2: Schematic diagram of the ZEUS analysis and MC reconstruction chain

- **beauty quarks from direct processes:**

(see BGF diagram in Fig. 2.2 left)

$$\gamma g \rightarrow b\bar{b};$$

- **beauty quarks from excitation in the photon:**

(see excitation diagram in Fig. 2.3 right)

$$bq \rightarrow bq;$$

$$bg \rightarrow bg;$$

- **beauty quarks from excitation in the proton :**

(see QCD Compton diagram in Fig. 2.2 right))

$$\gamma b \rightarrow bg;$$

$$gb \rightarrow bg;$$

$$qb \rightarrow bq;$$

$$b\bar{b} \rightarrow b\bar{b};$$

- **beauty quarks from hadron-like photon resolved processes:**

(see diagram in Fig. 2.3 left)

$$gg \rightarrow b\bar{b};$$

$$q\bar{q} \rightarrow b\bar{b};$$

In the event generation, the beauty quark mass was set to $m_b = 4.75 \text{ GeV}$ and the parton density functions used for the proton and the photon were CTEQ5L [24] and GRVG-LO [58] respectively. PYTHIA performs the fragmentation of the $b(\bar{b})$ quarks according to the Peterson fragmentation function with $\epsilon = 0.0041$. No preselection was applied.

5.3.2 Light flavour MC samples

In the case of light flavour (u, d, c) production, due to the low mass of the produced quarks, resolved photon processes are expected to be dominant. The light flavour MC samples generated are:

- **light flavour quarks from direct processes:**

$$q\gamma \rightarrow qg;$$

$$g\gamma \rightarrow q\bar{q};$$

- **light flavour quarks from resolved processes:**

$$q\bar{q} \rightarrow q\bar{q};$$

$$q\bar{q} \rightarrow q'\bar{q}';$$

$$q\bar{q} \rightarrow gg;$$

$$qq \rightarrow qq;$$

$$gg \rightarrow q\bar{q};$$

$$gg \rightarrow gg;$$

In this case, a pre-selection was applied in order to reduce the number of events to be analyzed, keeping only those of interest. The pre-selection consists in requiring two jets in the final state at true hadron level with the following characteristics:

- $E_T^{Jet1,2} \geq 4.0 \text{ GeV}$ and $-3 \leq \eta^{Jet1,2} \leq 3$

The proton PDF used was CTEQ5L and the minimal transverse momentum of the hard scatter, p_T^{min} , was set to 3 GeV .

Process	\mathcal{L} [pb^{-1}]	m_b [GeV]	Proton PDF	p_T^{min} [GeV]
bb direct	402.12	4.75	CTEQ5L	0
bb resolved	389.49	4.75	CTEQ5L	0
bb excitation in γ	409.67	4.75	CTEQ5L	0
bb excitation in p	440.42	4.75	CTEQ5L	0
inclusive lf. direct	187.60	\times	CTEQ5L	3
inclusive lf. resolved	81.19	\times	CTEQ5L	3

Table 5.1: Beauty quark mass, luminosity, proton PDF and minimal transverse momentum of hard scatter for the MC samples used in this work.

Chapter 6

Event Reconstruction

In this work, $b(\bar{b})$ quarks are identified by finding the decay vertices, commonly referred to as secondary vertices, of particles containing a $b(\bar{b})$ quark. The vertex reconstruction is therefore one of the most important ingredients of this analysis and requires a precise measurement of the particle trajectories and momenta.

This chapter starts describing how tracks are reconstructed from the data. Tracks of charged particles, as well as their momenta, are found using the CTD and the standard ZEUS track reconstruction software, as explained in section 6.1.

Moreover, an event containing a $b\bar{b}$ pair is characterized by the presence of at least 2 jets in the final state, coming from the hadronization of the $b(\bar{b})$ quarks. Tracking and energy information is combined to form *energy flow objects* (EFO's). Jets are made from EFO's using a clustering algorithm as outlined in section 6.3.

Finally, the reconstruction of the kinematic variables of interest is explained in section 6.4. The vertex reconstruction will be described in the following chapter.

6.1 Track Reconstruction

At ZEUS, tracks are reconstructed combining CTD hits to particle trajectories. Additionally information from other tracking detectors such as MVD, SRTD, RTD and FTD can be used. Moreover, the track reconstruction takes into account the errors on the hit measurements, the dead material distribution and the effects of multiple scattering. In principle, three different track types can be found according to the information used:

1. “CTD-only” tracks for which only the CTD information was required;

2. “Regular” tracks which combine both CTD and MVD information. Tracks are found in the MVD and linked to the tracks in the CTD using a combined pattern recognition;
3. “ZTT” tracks are “Regular” tracks in which the accuracy of the tracking close to the vertex has been improved with a technique known as Kalman filtering. This improves the tracking, taking into account effects like multiple scattering and kinks caused by the MVD material.

The software package used for track reconstruction in the first two cases is called VCTRACK [47] while ZTT tracks are computed by the package KFFIT [44]. For this analysis the ZTT tracks were used and thus its reconstruction will be explained here.

Given the coordinates of the hits in the CTD, MVD and STT detectors, the track and vertex reconstruction chain can be divided basically into four steps:

- Pattern recognition: hits are combined for a first estimation of the tracks (this step takes place within the VCTRACK package);
- Track fitting: the CTD hits are globally fitted by a helix;
- Kalman filtering procedure: the MVD hits are added to the fit including the effects of energy loss and multiple scattering;
- Vertex reconstruction: the vertices are identified and fitted.

6.1.1 Pattern recognition

The track pattern recognition starts at the outermost SuperLayer (SL9) in the CTD where the hit density is lower and follows trajectories inward through the MVD module layers. Two separate fits, both applying the least squares method, are used to model the trajectory of a track:

1. The “XY plane fit” is a circle fit with only 2 parameters (a_1, a_2), where a_1 is the angle tangent to the helix in the XY plane and $a_2 = Q/R$ (Q is the charge, R is the local radius). The third parameter usually required for a standard circle fit is fixed at $a_3 \equiv 0$ “pinning” each circle to a reference point (outermost axial CTD hit).
2. The “sZ plane fit” is a 2 parameter line fit. s is defined as the 2D path length such that $z = q_1 + sq_2$ where $q_1 = z_0$ at the reference point (x_0, y_0) and $q_2 = \cot \theta$. The reference point (x_0, y_0) for the trajectory

corresponds to the outer hit. The CTD z-by-timing system provides the z value with a resolution of approximately 5 *cm*.

6.1.2 Track fitting

The 2D trajectory fitting does not describe kinks due to coulomb scattering or energy-loss and contains neither drift velocity nor time of flight corrections. Thus, after the pattern recognition is done, a full 3D helix-parameter fit is performed in order to deal with these features providing high precision measurements.

For a magnetic field parallel to the beam axis (z axis in ZEUS geometry), which is the case of the magnetic field generated by the solenoid of the ZEUS detector to a good approximation, the trajectory of a particle can be parametrized as an axial helix $\vec{p} = (\phi_H, Q/R, QD_H, Z_H, \cot \theta_H)$, shown in Fig. 6.1, whose parameters are:

1. ϕ_H , the angle tangent to the helix in the XY plane,
2. Q/R , where Q is the track charge and R is local radius of curvature,
3. QD_H , where D_H is the distance of closest approach of the helix relative to the reference point,
4. Z_H , the z coordinate of the track at its closest approach to the reference point,
5. $\cot \theta_H$, where θ_H is the polar angle of the track.

The coordinates of the closest approach to the reference point can be written as:

$$X_H = X_{ref} + QD_H \sin \phi_H \quad (6.1)$$

$$Y_H = Y_{ref} - QD_H \cos \phi_H \quad (6.2)$$

$$Z_H. \quad (6.3)$$

The coordinates and the momentum of a generic point of the helix can then be parametrized as:

$$X = X_H + QR(-\sin \phi + \sin \phi_H) \quad (6.4)$$

$$Y = Y_H + QR(+\cos \phi - \cos \phi_H) \quad (6.5)$$

$$Z = Z_H + s(\phi) \cot \theta \quad (6.6)$$

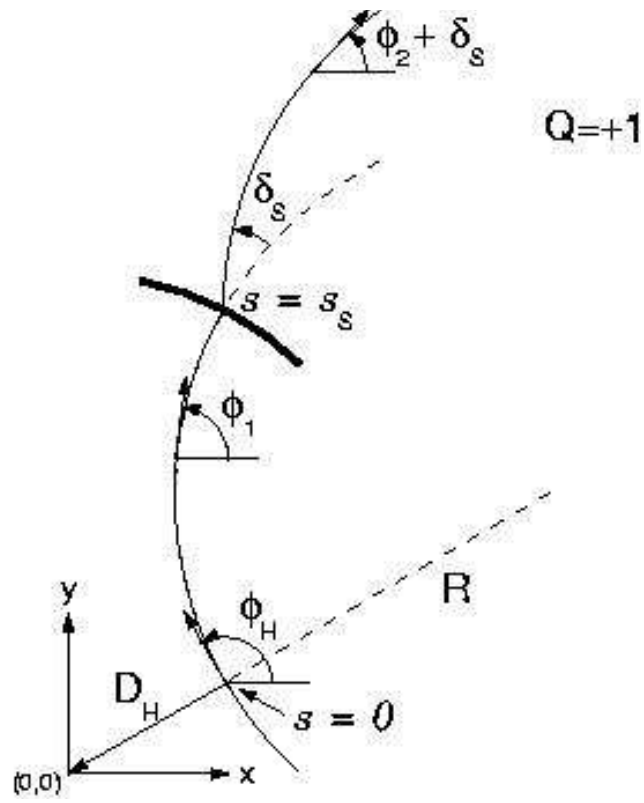


Figure 6.1: Helix parametrization in the XY plane of a charged track.

where $s(\phi) = -QR(\phi - \phi_H)$ is the path-length of a generic trajectory in the XY plane.

The track fit gives the best possible estimate of the five track parameters, based on the measurements and their uncertainties.

6.1.3 The Kalman filter track fit

The track parameters \vec{p} (and the 5×5 covariance matrix V associated to these parameters) obtained from the pattern recognition and track fitting steps are the seed information required by the Kalman filter to start. A complete description of the Kalman filter algorithm can be found e.g. in [44]. Here a brief review of the KF fit procedure applied to our specific case is given [60].

The Kalman filter is an efficient recursive data processing algorithm that incorporates all available information to estimate the current state of a dynamic system. Thus, its application to track fitting is straightforward if the track is interpreted as a discrete dynamic system [69]. The Kalman filter track fit algorithm consists of three basic steps:

- **Prediction:** estimation of the next state from the current state.
- **Filtering:** re-estimation of "present" state using the predicted state and including the current measurement.
- **Smoothing:** update of the previous state using the current filtered state.

Figure 6.2 shows graphically the 3 basic steps of the KF procedure: the Kalman filter starts with the first track hit at the detector surface $(i - 1)$ and makes a prediction for the hit at surface i . The prediction is combined with the actual measurement at surface i yielding an updated estimate. The process is repeated until the last detector surface is reached. During the smoothing step, the process is reversed and starts from the last hit. The track is then propagated backwards to the first detector layer using the complete available information.

The state of the track at surface i is given by the 5 parameters vector \vec{p}_i of the track at the intersection of the track with the surface i and the (5×5) covariance matrix V_i .

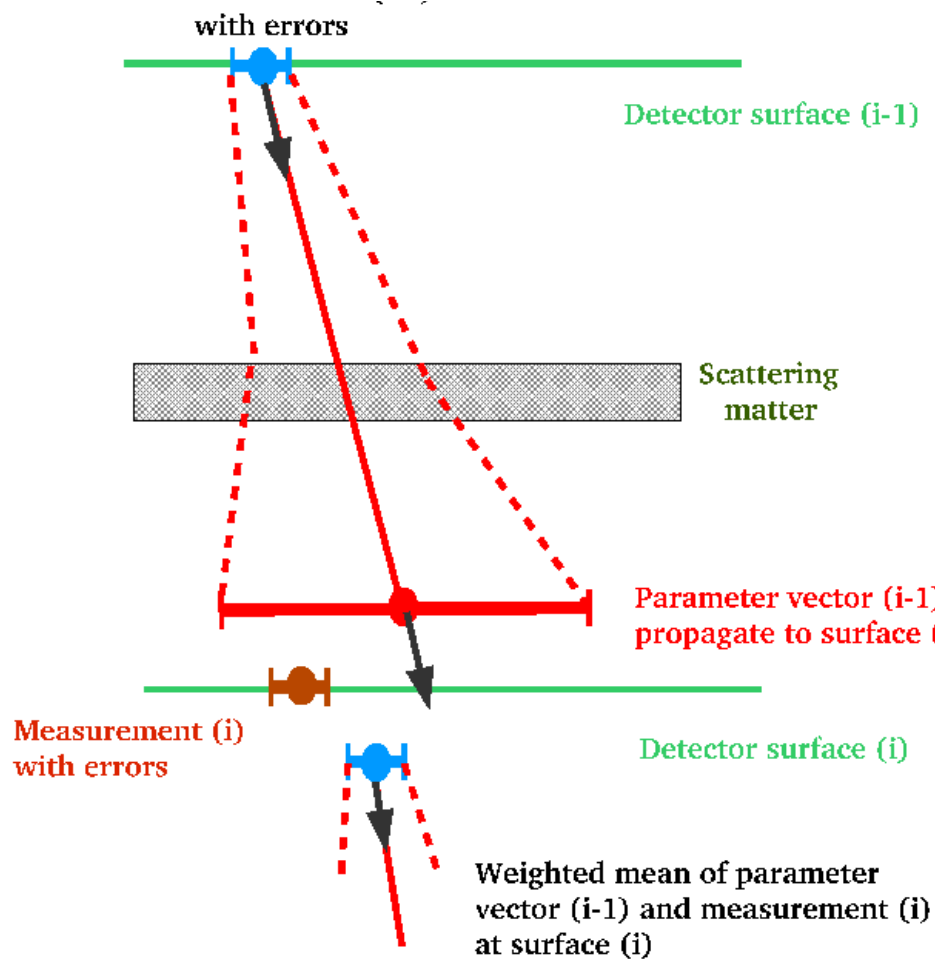


Figure 6.2: Example of a Kalman prediction step from detector (i-1) through scattering material to detector (i). The final state at detector (i) results from a filtering procedure by including the current measurement k .

The prediction step

In the prediction step the track is propagated from the initial state \vec{p}_{i-1} with covariance V_{i-1} to the next measurement plane (i.e. silicon sensor). Since the MVD layers are thin, the energy loss of the particles is neglected leaving unchanged the state vector. The track covariance increases due to multiple scattering. Thus, after propagation the state vector and the covariance are:

$$\vec{p}_i = \vec{p}_{i-1}, \quad (6.7)$$

$$V_i = V_{i-1} + M_i \quad (6.8)$$

where M_i is the contribution of the multiple scattering to the track's covariance between the detector surfaces $i-1$ and i . To keep the notation simple, the state and its covariance after the prediction step are defined as \vec{p}_{i-1} and V_{i-1} .

The filtering step

In the filtering step the trajectory of the particle is fitted to the current hit measurement i which has an uncertainty in the reconstructed position of σ_i . The weight of the hit is proportional to $1/\sigma_i^2$. The estimation of the state \vec{p}_i is a weighted mean of the prediction and the current measurement. For instance, if the measurement is very accurate and the hit has a very small error, its weight will be large and the trajectory is pulled more strongly towards this hit. The fit also depends on the track covariance. A track with large covariance is pulled more strongly than a track measured more precisely.

The state vector and its covariance after the filtering step are:

$$\vec{p}_i = \vec{p}_{i-1} + A_i(\vec{p}_i - \vec{p}_{i-1}) \quad (6.9)$$

$$V_i = V_{i-1} + A_i(V_i - V_{i-1})A_i^T \quad (6.10)$$

where A_i is the gain matrix $A_{i-1} = V_i V_{i-1}^{-1}$.

The χ^2 of the filtering step is the sum of two contributions: the difference Δy between the measured hit position and the predicted hit position on the silicon sensor (this quantity is a scalar in the case of a strip detector) and the difference between the predicted and filtered state vector $\Delta \vec{p}$. It can be written as:

$$\chi^2 = (\Delta y_i - A_i \Delta \vec{p})^T \frac{1}{\sigma_i^2} (\Delta y_i - A_i \Delta \vec{p}) + (\Delta \vec{p})^T V_i^{-1} (\Delta \vec{p}) \quad (6.11)$$

This χ^2 is used to decide whether to use the current hit to estimate the filtered vector state or move on to the next hit.

The smoothing step

The smoothing step consists in the estimation of the state vector at past times using the full available information. After the filtering step a filtered estimate \vec{p}_i and covariance matrix V_i have been calculated. If for the previous measurements all the parameters and covariances are stored, the final state vector estimate can be improved using the full information and propagating it back to all previous estimates:

$$\vec{p}_{i-1}^{smoothed} = \vec{p}_{i-1}^{filtered} - B_{i-1}(\vec{p}_{i-1}^{filtered} - \vec{p}_i), \quad (6.12)$$

$$V_{i-1}^{smoothed} = V_{i-1}^{filtered} - B_{i-1}(V_{i-1}^{filtered} - V_i)B_{i-1}^T, \quad (6.13)$$

$$B_{i-1} = V_{i-1}^{filtered}V_{i-1}^{-1} \quad (6.14)$$

When the effects of multiple scattering are large, B_{i-1} is small and the previous estimate of the trajectory has a larger weight than the current estimate so that its influence on the smoothing procedure is small.

6.2 Reconstruction of hadronic system

The reconstruction of the hadronic final state is important for the identification of events with $b\bar{b}$ pairs since 2 jets are demanded. No electron or muon is required (inclusive analysis). Only the hadronic system will be present in the reconstructed data set of this analysis. Also the different kinematic variables which define the event properties are calculated from the hadronic final state.

To get the most precise energy determination, information from the calorimeter (CAL) and the tracking detector is combined. The energy resolution of the CAL is $\sigma(E)/E \sim 1/\sqrt{E}$ which increases for higher particle energies while the tracking system resolution, given by $\sigma(p_T)/p_T \sim a \cdot p_T \oplus b \oplus c \cdot p_T$, provides a better energy estimation for lower particle energies (see Fig. 6.3). As it can be seen in Fig. 6.3, the tracking information is mainly used below 10 GeV while the calorimetric information is used above it.

The combination of these two pieces of information gives the energy-momentum four vectors known as EFOs, representing the oriented energy deposition

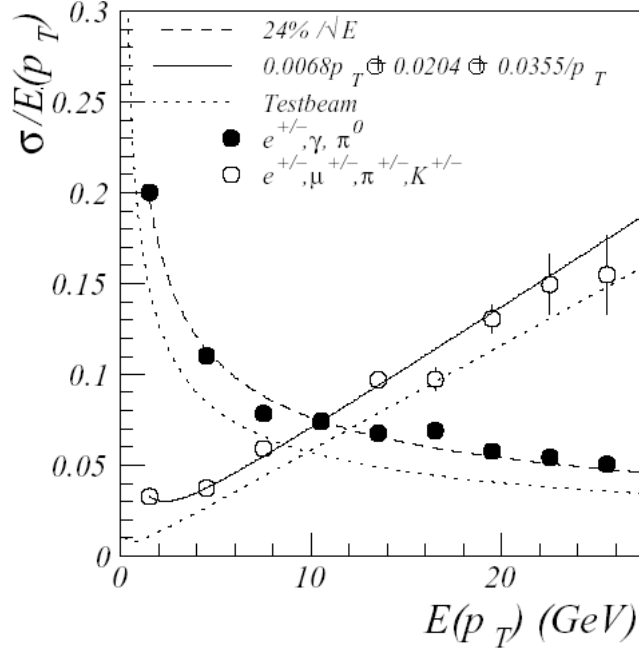


Figure 6.3: Resolution of single particles from MC simulations [8]. The track transverse momentum resolution in the CTD (open markers) and the CAL energy resolution (closed markers) are shown

of particles traveling in the detector. For neutral particles only CAL information is used.

6.2.1 Reconstruction of Energy Flow Objects

The reconstruction of EFOs [8] is done in several steps due in part to the geometry of the ZEUS calorimeter (section 4.2.3) and to the combination of CAL and CTD information. All calorimetric variables are constructed from the basic measurement, which is the energy content of a cell. In the first step, *cell islands* are found. Cell islands are two dimensional clusters of neighboring cells in the three separate layers of the CAL(EMC, HAD1, HAD2) and are formed grouping all neighboring cells with an energy bigger than a threshold, starting from the cell with the highest energy.

In the second step, the islands belonging to a shower of a single particle or a jet of particles are grouped into the so called *cone islands* (3-dimensional objects) using the angular separation of the cell islands. The matching starts at the outermost CAL layer (HAC2) and proceeds inward guided by single pion MC probabilities [8]. The cone island center is calculated by the logarithmic

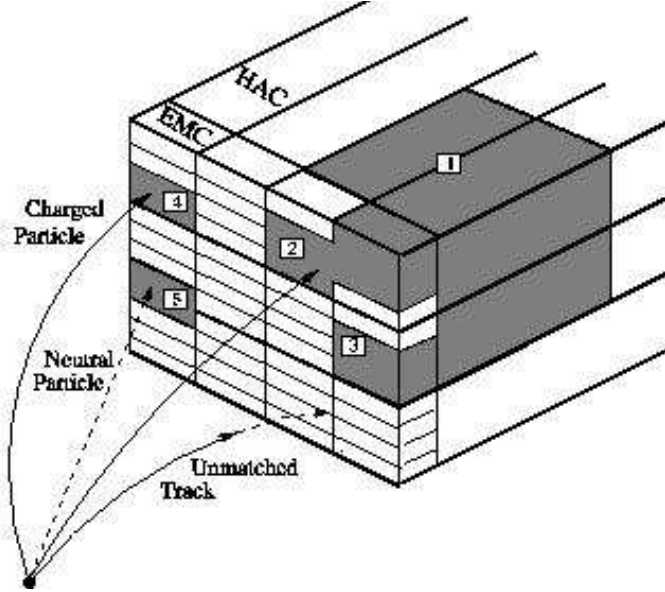


Figure 6.4: Schematic plot of the island clustering algorithm. Neighboring calorimeter cells are clustered into cell islands. Four EMC and one HAC1 island are shown. The island 1,2 and 3 are matched to form a cone island. The island 4 and the cone island (1,2,3) are matched with tracks (charged particles). No track is matched to island 5 (neutral particle). An unmatched track corresponding to a low momentum particle is also shown.

center-of-gravity of the shower.

After the second step, the cone island is matched with the charged CTD tracks. First, the tracks must be extrapolated to the surface of the CAL taking into account the magnetic field geometry and further into the CAL by a linear approximation using the track momentum vector at the CAL surface. The tracks must fulfill certain requirements: they must be fitted to the primary vertex and must have crossed at least 4 CTD superlayers with a transverse momentum in the range $0.1 < p_T < 20 \text{ GeV}$. If the track has hits in more than 7 superlayers, the upper limit is increased to $p_T < 25 \text{ GeV}$.

The matching of a track to a cone island uses the Distance of Closest Approach method (DCA). The match takes place if the distance is smaller than the cone island radius or less than $r_{min} = 20 \text{ cm}$. Figure 6.4 illustrates how the clustering to cone islands takes place as well as the track-island matching.

The EFO momentum is taken from either the CAL cluster or the matched CTD track depending on the resolution of the track momentum and the CAL cluster energy resolution. Charged tracks with no associated calorimetry

object are assumed to be low energy pions and the CTD information is used for them. Unmatched energy deposits are considered to come from neutral particles and the CAL energy measurement is used. Cone islands with more than 3 matched tracks are treated as jets and the CAL information is used.

6.2.2 EFO and cone island energy correction

Discrepancies between data and MC simulations in the reconstruction of EFOs have been found in some studies [86]. Some of the causes of these discrepancies and the corresponding corrections are listed below:

- Corrections for energy loss in dead material: this kind of energy loss is generally difficult to implement fully in the MC simulations. A detailed map of the dead material distribution is shown in figure 6.5 and is used to write correction algorithms. The main contributions to energy loss are from the beam pipe, the tracking detector and the solenoid.
- Corrections for the CAL geometry not properly described in MC: the geometry of the cracks between the calorimeter sections is not well described in MC and needs to be corrected.
- Corrections for energy overestimation of low momentum hadrons: at low energy (below $\sim 1 \text{ GeV}$) the energy loss due to ionization dominates over the losses in hadronic interactions reducing the uranium compensation effect ($e/h \sim 0.6$). This effect causes an overestimation of the particle energy.

6.3 Jet Reconstruction

One of the experimental signatures of an event containing a $b\bar{b}$ pair is at least 2 jets in the final state. A jet is a collection of approximately collinear colorless hadrons originating from the hadronization of the colored partons, in our case b and \bar{b} quarks and gluons (excitation processes). Thus, the reconstructed final state jets, correlated with the partons from which they originated, provide a kinematical signature of the underlying dynamics.

However, the exact definition of a jet is not unique and depends on the algorithms used to reconstruct it. In an attempt to standardize the jet definition, some criteria were established during the Snowmass workshop in 1990 [26]. In particular, two requirements must be fulfilled by every jet finder:

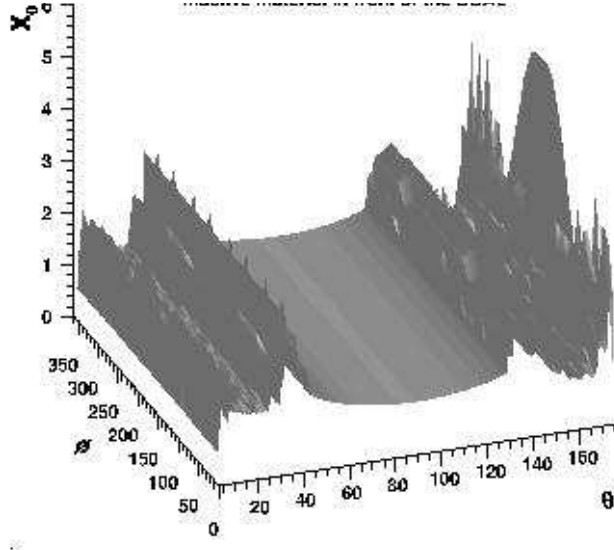


Figure 6.5: Map of dead material distribution between the interaction region and the CAL in units of the radiation length X_0 in the $\theta - \phi$ plane.

1. Infrared safety: the results should be independent of the emission of very low energy particles.
2. Collinear safety: the results should be independent of one parton splitting into two partons moving collinearly.

In addition to this, a jet algorithm suited for ep collision must fulfill two further requirements:

3. correct treatment of the proton and, if present, photon remnants, and
4. independence from longitudinal Lorentz transformations (commonly interesting frames are Lorentz boosted along the beam direction with respect to laboratory frame).

In this analysis, the jets were reconstructed using a k_T -type clustering algorithm [10] based on EFO's (see section 6.2.1). The resolution variable used by this algorithm is the relative transverse momentum p_T between particles. The algorithm distinguishes between two classes of final state jets which originate from the hard (macro jets) and soft (remnant jets) interactions. Thus a hard scattering scale is used to separate hard and soft processes.

The KTCLUS algorithm clusters the final state objects (EFOs) into jets according to the following steps:

1. For every EFO e_k and for every pair e_k and e_l , the resolutions variables d_k and d_{kl} are defined as:

$$\begin{aligned} d_k &= E_k^2 \theta_k^2 \\ &\simeq p_{T,k}^2 \text{ for } \theta \rightarrow 0 \end{aligned} \quad (6.15)$$

$$\begin{aligned} d_{kl} &= \min(E_k^2, E_l^2) \theta_{kl}^2 \\ &\simeq p_{T,kl}^2 \text{ for } \theta \rightarrow 0 \end{aligned} \quad (6.16)$$

2. Scale the d_{kl} by ΔR_{kl}^2 , where $\Delta R_{kl}^2 = (\eta_k - \eta_l)^2 + (\phi_k - \phi_l)^2$.

3. Find the smallest value among d_k and d_{kl} : $d_{min} = \min(d_k, d_{kl})$

- If $d_{min} = d_{ij}$, the two EFOs are merged into a single new *pseudoparticle* using the *E recombination scheme* where the hadrons are treated as massive particles and their four-momenta are added as: $p_{ij} = p_i + p_j$. The new resolution variables are given by:

$$d_{(ij)} = (p_{T,i} + p_{T,j})^2 \quad (6.17)$$

$$d_{(ij)k} = \min((p_{T,i} + p_{T,j})^2, (p_{T,k})^2) \Delta R_{(ij)k}^2 \quad (6.18)$$

- If $d_{min} = d_k$, the EFO e_k is removed from the EFO list and included in a so called remnant jet.
4. This procedure is repeated until all the EFOs or pseudoparticles not included in the remnant jets have resolution variables larger than the hard scale d_{cut} where $\lambda^2 \ll d_{cut} \ll s$ with λ the QCD scale and \sqrt{s} the center-of-mass energy.

This procedure results in *hard final-state* jets and remnant jets. The E recombination scheme is favoured over other schemes because the massive approach provides the best agreement between the jets of heavy flavour beauty quark measured in the detector and their theoretical predictions [61].

6.4 Reconstruction of kinematic variables

As we will see in chapter 8, kinematic variables are used to distinguish between the photoproduction (PHP) and the deep inelastic scattering (DIS) regimes as well as to remove events not of interest to this analysis. There are

many methods to reconstruct the inelasticity y . The two used in this thesis are the *electron method* and the *Jacquet-Blondel method* and both of which are described here.

The *electron method* can be used when the kinematics of the outgoing lepton is known with precision, that is the energy of the scattered lepton E'_e as well as the scattering angle θ_e (CAL measurements). Given an initial energy of lepton E_e , the inelasticity y can be expressed as:

$$y_e = 1 - \frac{E'_e}{2E_e}(1 - \cos \theta_e) \quad (6.19)$$

In photoproduction events the outgoing lepton can not be detected since it is scattered at very low angles and escapes into the beam pipe. At ZEUS, the maximum polar angle at which the lepton can be identified is $\theta_e = 177^\circ$. In this case we can use the *Jacquet-Blondel method* which reconstructs the variables y from the hadronic final state as:

$$y_{JB} = \frac{\sum_f (E_f - p_{zf})}{2E_e} \quad (6.20)$$

where E_f are the energies and p_{zf} the momenta parallel to the beam pipe and f runs over all the final state particles (EFO's, see section 6.2.1). The quantity $\sum_f (E_f - p_{zf})$ is conserved within good approximation if all the final state particles are measured correctly in the detector,

$$\sum_f (E_f - p_{zf}) = \sum_f E_f(1 - \cos \theta_f) \approx \sum_i E_i(1 - \cos \theta_i) = 2E_e \quad (6.21)$$

where i runs over the initial state particles, the electron(positron) and proton. For DIS events, the scattered lepton contributes to the measurement of this quantity even if it is not identified, wrongly yielding $y_{JB} \sim 1$. Therefore, a cut on y_{JB} will be used to remove residual DIS events (see chapter 8 for event selection)

Chapter 7

Vertex Reconstruction

This chapter describes the algorithms for the reconstruction of the primary and secondary vertices, as performed for the first part of this analysis. The vertex reconstruction consists of two main stages: vertex finding and fitting. Vertex finding involves the identification of the tracks belonging to a same decay vertex while the vertex fitting implies the estimation of the vertex position as well as the track parameters at the vertex.

As mentioned in the previous chapter, VCTRACK is the traditional ZEUS software package responsible for the track and vertex reconstruction. However, the highest priority of the VCTRACK vertex package is to find the primary vertex while secondary vertices are found only to enhance the primary vertex finding. This, of course, is not enough for an analysis which aims to tag beauty events by identifying decay vertices of B-hadrons. Thus the reconstruction technique for secondary vertices has been redesigned to fulfill the requirements of this analysis i.e. to provide more accurate secondary vertices positions and with a higher probability to belong to B hadrons. Secondary vertices are then reconstructed using the user **VerteX Utility** (VXU) algorithm [43] while the VCTRACK package has been used only for the determination of the primary vertex. Nevertheless, the vertex fit (mathematical formalism) performed by VCTRACK and the one by VXU are both the same, based on a fast vertex fitting method using a local parametrization of tracks [67] and therefore it will be described next.

This chapter is organized as follows. First, a brief overview of the whole procedure is given in order to get a clear idea of what steps are involved in the process. Then, section 7.1 will describe the general vertex fitting algorithm on which the determination of the primary and secondary vertex is based. A complete description of the mathematics can be found in [67]. To achieve a higher accuracy, the primary vertex will be replaced by an average

value calculated on an event-by-event basis as explained in section 7.3. The secondary vertex reconstruction will be explained in section 7.4.

Overview of the steering procedure

The highest priority of the VCTRACK vertex package is to find the primary vertex. Secondary vertices are found only to enhance the primary vertex finding. The algorithm takes the reconstructed tracks from the track fitting phase as the starting point and performs the following steps:

1. Preliminary filtering: the trajectories which are not compatible with the beamline are removed;
2. Vertex simple fit: a “simple” and fast algorithm which provides the vertex vector $\mathbf{V} = (V_x, V_y, V_z)$. After \mathbf{V} has been obtained, the vertex χ^2 is calculated by adding the contributions $\Delta\chi_i^2$ of each participating track i . Tracks which lead to an unacceptably large χ^2 (based on tuned cuts) are discarded and the vertex is re-estimated. This procedure is repeated until a good χ^2 is obtained;
3. Vertex full fit: the final vertex position and the refitted trajectory for each track are found by an algorithm that simultaneously finds the vertex position and re-adjusts the track parameters at the vertex;
4. Secondary vertices finding: if the VCTRACK package is run in “multi-vertex mode” (also “primary-vertex only” mode possible), compatible secondary vertices are identified after the primary vertex by pairing all reconstructed tracks which have not been refitted to the primary vertex.

7.1 The “full” vertex fitting algorithm

The input information for the vertex fit are the 5-parameter vectors \mathbf{q}_{ij} ($i = 1 \sim n, j = 1 \sim 5$), with n the number of tracks, and their covariance matrices \mathbf{C}_i . For a track i , the vector \mathbf{q}_{ij} describes the trajectory of the charged particle close to the origin point which is a first approximation of the primary vertex or the secondary vertices. Since the goal is to determine the vertex position and the track momenta at the vertex, \mathbf{q}_{ij} will be expressed as a function of the vertex position $\mathbf{V} \equiv (x_v, y_v, z_v)$ and the track momenta \mathbf{p}_{im} ($m = 1 \sim 3$) at the vertex: $\mathbf{q}_{ij} = F_j(V, \mathbf{p}_i)$ (see Fig. 7.1). The goal is to find the \mathbf{V} and \mathbf{p}_i which minimize χ^2 :

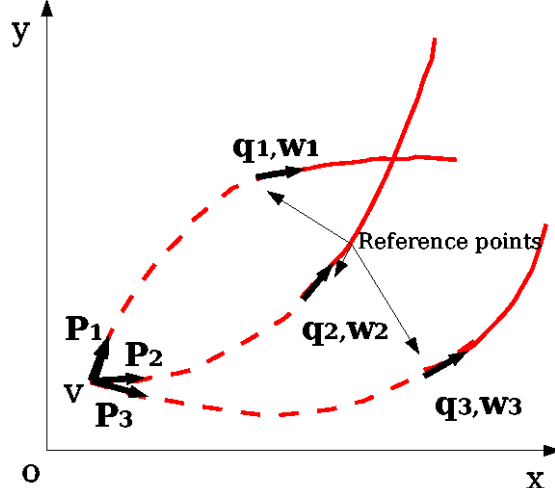


Figure 7.1: Relation between \mathbf{q}_i and (V, p_i) in a vertex fit.

$$\chi^2 = \sum_i \Delta \mathbf{q}_i^T \mathbf{W}_i \Delta \mathbf{q}_i, \quad (7.1)$$

where

$$\Delta \mathbf{q}_i = \mathbf{q}_i^{measured} - \mathbf{F}(V, p_i),$$

$\Delta \mathbf{W}_i = \mathbf{C}_i^{-1}$ is the weight matrix.

For small variations δV and δp_i around first approximations \mathbf{V}_0 and \mathbf{p}_i^0 , it is possible to linearize the functions \mathbf{F} as:

$$\mathbf{F}(V^0 + \delta V, p_i^0 + \delta p_i) = \mathbf{F}(V^0, p_i^0) + D_i \delta V + E_i \delta p_i, \quad (7.2)$$

where the matrices \mathbf{D}_i and \mathbf{E}_i are expressed as:

$$(\mathbf{D}_i)_{jn} \equiv \frac{\partial \mathbf{F}_j(\mathbf{V}, \mathbf{p}_i)}{\partial V_n}, \quad (7.3)$$

$$(\mathbf{E}_i)_{jm} \equiv \frac{\partial \mathbf{F}_j(\mathbf{V}, \mathbf{p}_i)}{\partial p_{im}}, \quad (7.4)$$

with $j = 1 \sim 5, n = 1 \sim 3, m = 1 \sim 3$.

In this way, substituting equation 7.2 in 7.1, the χ^2 can be expressed as:

$$\chi^2 = \sum_i (\delta \mathbf{q}_i - D_i \delta V - E_i \delta p_i)^T \times \mathbf{W}_i (\delta \mathbf{q}_i - D_i \delta V - E_i \delta p_i) \quad (7.5)$$

Minimizing the χ^2 with respect to \mathbf{V} :

$$\partial\chi^2/\partial\mathbf{V} = 0 \leftrightarrow (\sum_i \mathbf{D}_i^T W_i D_i) \delta\mathbf{V} + (\sum_i \mathbf{D}_i^T W_i E_i) \delta\mathbf{p}_i = \sum_i \mathbf{D}_i^T W_i D_i \delta q_i$$

and with respect to \mathbf{p}_i :

$$\partial\chi^2/\partial\mathbf{p}_i = 0 \leftrightarrow (\mathbf{E}_i^T W_i D_i) \delta\mathbf{V} + (\mathbf{E}_i^T W_i E_i) \delta\mathbf{p}_i = \mathbf{E}_i^T W_i \delta q_i$$

both equations can be rewritten as:

$$\mathbf{A} \delta V + \sum_i \mathbf{B}_i \delta p_i = \mathbf{T} \quad (7.6)$$

$$\mathbf{B}_i^T \delta V + \sum_i \mathbf{C}_i \delta p_i = \mathbf{U}_i \quad (7.7)$$

where

$$\begin{aligned} \mathbf{A} &\equiv \sum_i \mathbf{D}_i^T W_i D_i, \mathbf{B}_i \equiv \mathbf{D}_i^T W_i E_i, \mathbf{C}_i \equiv \mathbf{E}_i^T W_i E_i, \\ \mathbf{T} &\equiv \sum_i \mathbf{D}_i^T W_i \delta q_i, \mathbf{U}_i \equiv \mathbf{E}_i^T W_i \delta q_i. \end{aligned}$$

The solution for the vertex position is then:

$$\mathbf{V} = \mathbf{V}^0 + \delta V = \mathbf{V}^0 + (\mathbf{A} - \sum_i B_i C_i^1 B_i^T)^{-1} (\mathbf{T} - \sum_i B_i C_i^1 U_i) \quad (7.8)$$

The term $(\mathbf{A} - \sum_i B_i C_i^1 B_i^T)^{-1}$ is just the covariant matrix $Cov(V, V)$ associated to $\delta\mathbf{V}$. Substituting $\delta\mathbf{V}$ back into equation 7.7, $\delta\mathbf{p}_i$, $Cov(\mathbf{p}_i, p_j)$ and $Cov(\mathbf{V}, p_i)$ can be obtained.

There are several advantages in using this method instead of the standard least-squares method. First of all, the computing time is approximately proportional to the number of tracks n while for the standard least squares method, consisting in the inversion of $(3n + 3) \times (3n + 3)$ matrices, it is proportional to n^3 . Hence this method is much faster and particularly suitable to be applied in cases of high track multiplicity.

Another advantage relies on the fact that the matrices $\mathbf{A}, \mathbf{B}_i, \mathbf{T}$ and \mathbf{U}_i are kept in memory so that to add/remove a track to/from the fitted vertex is easier and does not need a complete recalculation from the beginning [66].

Moreover, within this method the information of the beam position and profile can be used for the estimation of the primary vertex. If $\mathbf{b} = (x_b, y_b, z_b)$ are the coordinates of the beam spot and \mathbf{C}_b is its covariance matrix, one just needs to add $(V - b)^T C_b^{-1} (V - b)$ to the χ^2 in eq. 7.1.

7.1.1 The "perigee" parametrization

The aim of this parametrization is to describe in a precise and simple way the particle trajectory in the neighborhood of the expected vertex. The extrapolation of the trajectory to the vertex region and its error matrix is performed once per track.

The magnetic field is assumed to be along the z-axis (beam direction). The "perigee" P is the point of closest approach of the trajectory to the z-axis and is defined through 5 parameters $q \equiv (\epsilon, z_p, \theta, \phi_p, \rho)$ (see Fig. 7.2):

- ϵ is the algebraic value $\epsilon \equiv OP$ with O being the origin of the coordinate system (if the origin O is chosen to be around the interaction point, this point will be close to the primary vertex and short-lived secondary vertices),
- the coordinate z_p ,
- the polar angle θ of the trajectory with respect to the z-axis,
- and its signed curvature ρ ($1/|\rho|$ is the radius of curvature in the x-y projection and the sign is positive if the trajectory is anticlockwise)

With these conventions, the trajectory around P can be parametrized like:

$$x \approx \epsilon \sin \phi_p + L \cos \phi_p - \frac{L^2 \rho}{2} \sin \phi_p, \quad (7.9)$$

$$y \approx -\epsilon \cos \phi_p + L \sin \phi_p + \frac{L^2 \rho}{2} \cos \phi_p, \quad (7.10)$$

$$z \approx z_p + L \cot \theta \quad (7.11)$$

where L is the distance from P along the trajectory in the $x - y$ plane.

As stated in the section above, to use the perigee parametrization in the vertex fit the parameters must be expressed as functions of the vertex position $\mathbf{V} \equiv (x_v, y_v, z_v)$ and the track parameters $\mathbf{p} = (\theta, \phi_v, \rho)$ at the vertex. The parameters θ and ρ do not change if we pass from \mathbf{V} to P. Introducing the

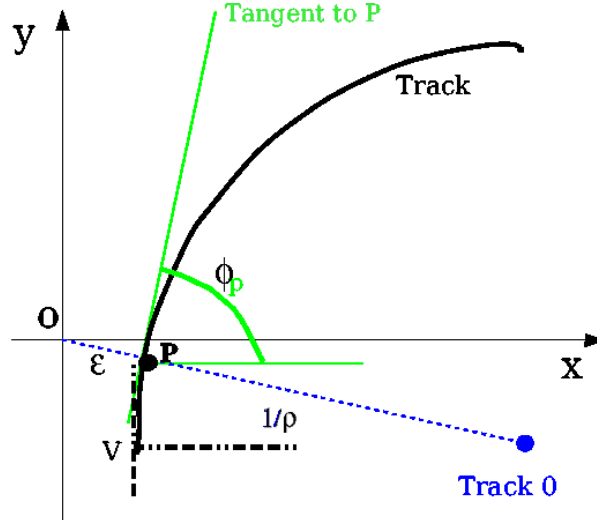


Figure 7.2: Perigee parameters in the $x - y$ projection, where ϵ is the distance between P and O, z_p is the z coordinate of P, θ is the polar angle of the trajectory w.r.t. the z-axis at P, ϕ_p is the angle of the trajectory at P and $1/\rho$ is the radius of the track curvature.

quantities $Q \equiv x_v \cos \phi_v + y_v \sin \phi_v$ and $R \equiv y_v \cos \phi_v - x_v \sin \phi_v$, calculations at first order in ρ give:

$$\epsilon = -R - Q^2 \rho / 2 \quad (7.12)$$

$$z_p = z_v - Q(1 - R\rho) \cot \theta \quad (7.13)$$

$$\phi_p = \phi_v - Q\rho \quad (7.14)$$

which is used to calculate the matrices of derivatives 7.3 and 7.4:

$\partial \downarrow / \partial \rightarrow$	x_v	y_v	z_v	θ	ϕ_v	ρ
ϵ	s	-c	0	0	Q	$-Q^2/2$
z_p	-tc	-ts	1	$+Q(1+t^2)$	-Rt	QRt
ϕ_p	$-\rho c$	$-\rho s$	0	0	1	-Q
	D matrix			E matrix		

Table 7.1: D and E matrices components

where $c \equiv \cos \phi_v$, $s \equiv \sin \phi_v$ and $t \equiv \cot \theta$.

7.2 The "simple" vertex fitting

Using the perigee parametrization allows the usage of an easier algorithm in which the vertex position can be estimated without re-fitting the track parameters. In this algorithm, the variation of the transverse errors along the tracks is neglected in the neighbourhood of the perigee and $\mathbf{p} = (\theta, \phi_v, \rho)$ is considered constant. Therefore, the 5×5 covariance matrix \mathbf{C} (eq. 7.1) can be reduced to the 2×2 submatrix \mathbf{C}' corresponding to the remaining variables z_v and ϵ . On the same way, the 5×3 matrix \mathbf{D} is reduced to its 2×3 submatrix \mathbf{D}' (derivatives of (ϵ, z_v) w.r.t. (x_v, y_v, z_v)).

If \mathbf{W}' is the inverse of \mathbf{C}' and $q' = (\epsilon, z_p)$, the equations 7.2 and 7.5 can be reduced to:

$$\mathbf{F}(V^0 + \delta V, p_i^0 + \delta p_i) = F(V^0, p_i^0) + D'_i \delta V, \quad (7.15)$$

$$\chi^2 = \sum_i (\delta \mathbf{q}'_i - D'_i \delta V)^T \times \mathbf{W}'_i (\delta \mathbf{q}'_i - D'_i \delta V) \quad (7.16)$$

and following the same procedure than before, the solution for the vertex position will be:

$$\mathbf{V} = \mathbf{V}^0 + \delta V = \mathbf{V}^0 + \left(\sum_i D_i'^T W'_i D'_i \right)^{-1} \left(\sum_i D_i'^T W'_i \delta q'_i \right) = A^{-1} T. \quad (7.17)$$

If \mathbf{V}^0 is chosen conveniently as the origin, $\delta \mathbf{V} = V^0$ and $\delta \mathbf{q}'_i = q'_i$. Let $\mathbf{x}_{pi} = (x_{pi}, y_{pi}, z_{pi})$ be the coordinates of the perigee P of the i^{th} track, then we will have:

$$x_{pi} = \epsilon \sin \phi_p + L \cos \phi_{pi} \quad (7.18)$$

$$y_{pi} = -\epsilon \cos \phi_{pi} \text{ or } x_{pi} = \mathbf{D}_i'^{-1} q'_i \quad (7.19)$$

$$z_{pi} = z_{pi} \quad (7.20)$$

and χ^2 and the vertex \mathbf{V} can be rewritten as:

$$\chi^2 = \sum_i (x_{pi} - \mathbf{V})^T \mathbf{w}_i \sum_i (x_{pi} - \mathbf{V}); \quad (7.21)$$

$$\mathbf{V} = \left(\sum_i \mathbf{w}_i \right)^{-1} \left(\sum_i \mathbf{w}_i x_{pi} \right) \quad (7.22)$$

with $\mathbf{w}_i = D_i'^T W'_i D'_i$. The \mathbf{V} error matrix is simply $(\sum_i \mathbf{w}_i)^{-1}$.

This "simple" method provides a fast way to determine if a track is compatible with the vertex or not. When solving \mathbf{V} by eq. 7.22, the two factors on the right can be stored. Then, the contribution to the χ^2 (eq. 7.21) by every single track can be considered. If a track k would contribute too much to the χ^2 , the vertex can be re-estimated by subtracting the corresponding \mathbf{w}_k and $\mathbf{w}_k x_{pk}$ from the equation 7.22, a complete calculation from the beginning not being necessary. Again, the information of the beam spot can be used by adding the corresponding position information and weight matrices to the equation 7.22.

7.3 Beam spot

The analysis presented in this thesis is based on the measurement of the decay length of the particles. This measurement depends directly on the position of the primary and secondary vertices. Hence, an accurate determination of the primary vertex is essential.

The estimation of the primary vertex position is based on the fast vertex fitting algorithm using the perigee parametrization of the tracks as described in the previous sections. However, if the primary vertex spread¹ in the transverse plane $x - y$ is smaller than its reconstruction resolution, a higher precision on its position can be achieved by replacing its value in each event with its average obtained from the vertex position over all the events in the same data taking period. This average is technically named *beam spot*.

Some cuts were applied on calorimetric variables to reduce the background due to beam-gas interactions. Only tracks with a minimal quality were selected to be fitted. A detailed description of the cuts and the procedure can be found in reference [77] and reference [75].

During the 2004 data taking the beam spot shifted its position [87] due to the movement of the GO/GG magnets (between fills), or temperature effects, luminosity, background and polarization tuning (during fills) (see Fig.7.3). This leads to two kind of dependences of the beam spot:

- **Time dependence:** due to the drift of the beam spot position during a fill. However, these time variations were found to be small compared to the sigmas of the beam spot coordinate distributions and no correction was applied.

¹The expected transverse widths of the HERA beams at the interaction point are $\sigma_x = 110 \mu m$ and $\sigma_y = 30 \mu m$.

- **z dependence:** due to the fact that the detector and beam axes are not parallel along the longitudinal direction which gives rise to a dependence of the primary vertex X, Y coordinates on the Z coordinate. Figure 7.4 shows how this effect can cause a variation of up to $\sim \pm 200 \mu m$ which is not negligible for this analysis and must be corrected. The new coordinate positions can be then written as

$$X_{BS}^{corrected} = X_{BS} + a1_x(Z_{primaryvtx} - Z_{BS}) \quad (7.23)$$

$$Y_{BS}^{corrected} = Y_{BS} + a1_y(Z_{primaryvtx} - Z_{BS}) \quad (7.24)$$

where $a1_x$ and $a1_y$ are the parameters obtained after fitting the distributions of Fig. 7.4 with a straight line.

7.4 Secondary Vertex Reconstruction

In contrast to the standard reconstruction package VCTACK, VXU allows the user to select the tracks used as input into the algorithm. Hence, only selected tracks are fitted to a vertex. Moreover, tracks which lead to a bad χ^2 of the vertex ($\chi^2 > 100$) are discarded automatically.

If the algorithm finds a vertex, the fit procedure outputs the three vertex coordinates, the covariance matrix and the χ^2 . Fig. 7.5 shows the secondary vertex resolution. It is performed comparing the reconstructed vertex with the Monte Carlo generated vertex for a beauty photoproduction sample. Secondary interactions and photon conversions in the beampipe or in the MVD material are rejected by removing all vertices which lie at more than 1 cm in the $x - y$ plane and 30 cm in the z plane from the nominal interaction point². If two oppositely charged tracks come from the same secondary vertex and yield an invariant mass compatible with K_s^0 or Λ , those tracks are removed from the group of tracks selected to perform the vertex. The distributions were parametrized with a Breit-Wigner distribution. The typical resolutions obtained were about $260 \mu m$ FWHM which corresponds to a $\sigma \sim 110 \mu m$ ³ at 68% confidential level. These values are consistent with expectations [45].

² $v_x = 1.92 \text{ cm}, v_y = 0.17 \text{ cm}, v_z = 0.0 \text{ cm}$

³ $\sigma = FWHM/2.35$

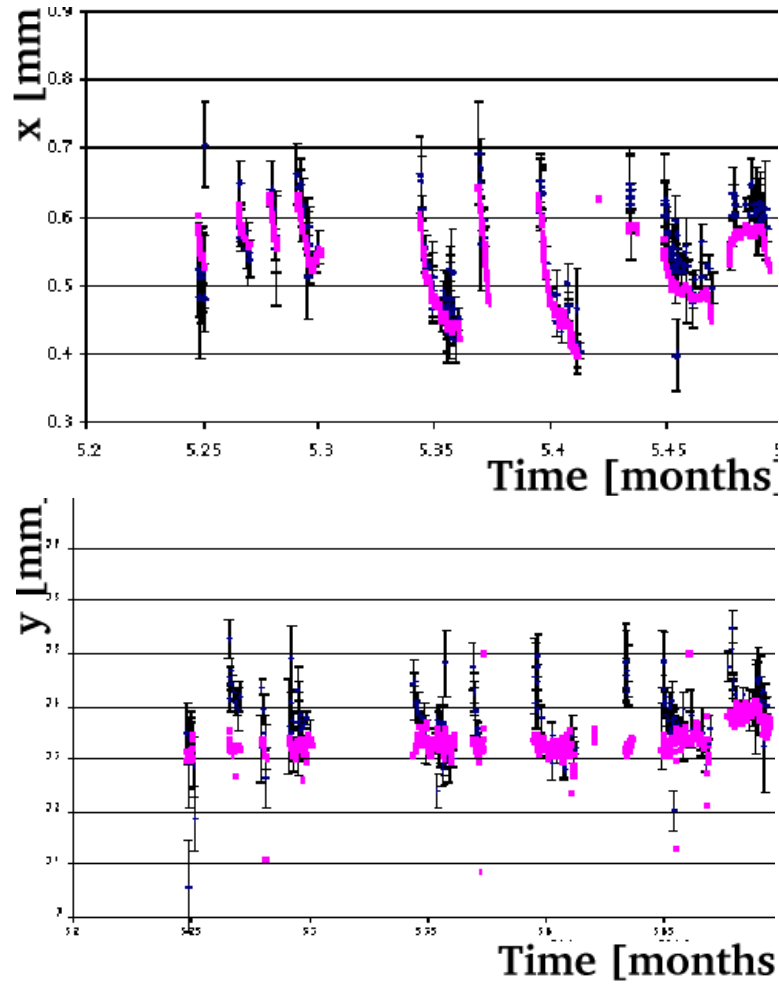


Figure 7.3: HERA(lilac) and H1(black) vertical and horizontal beam position measurements during the 2004 data taking period. The periodic structure of the plot represents the different machine fills.

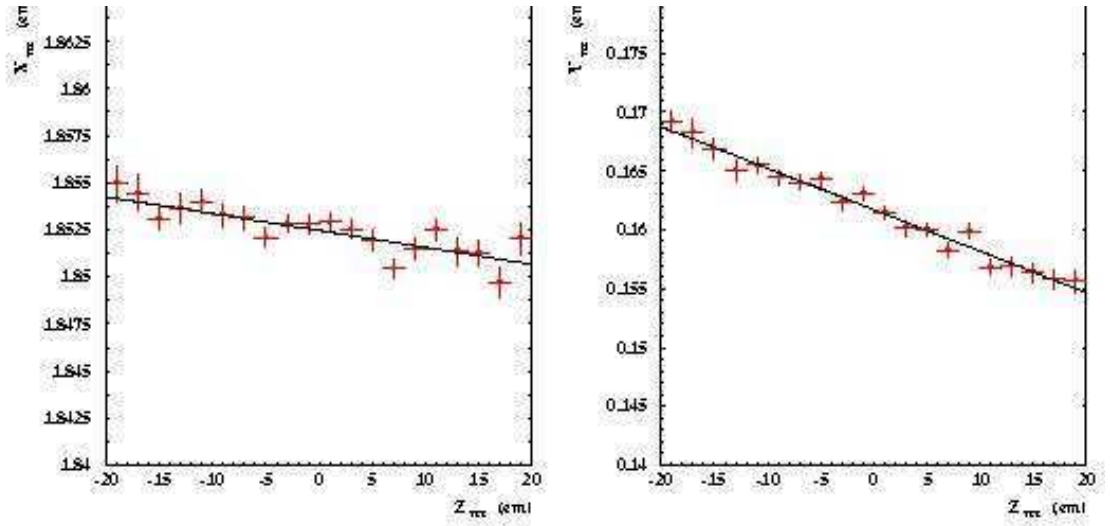


Figure 7.4: Dependence on Z of the primary vertex coordinates X and Y for 2004 data. The distribution was fitted with a straight line. Only the central region around the mean value of Z is shown ($|Z| < 20\text{cm}$)

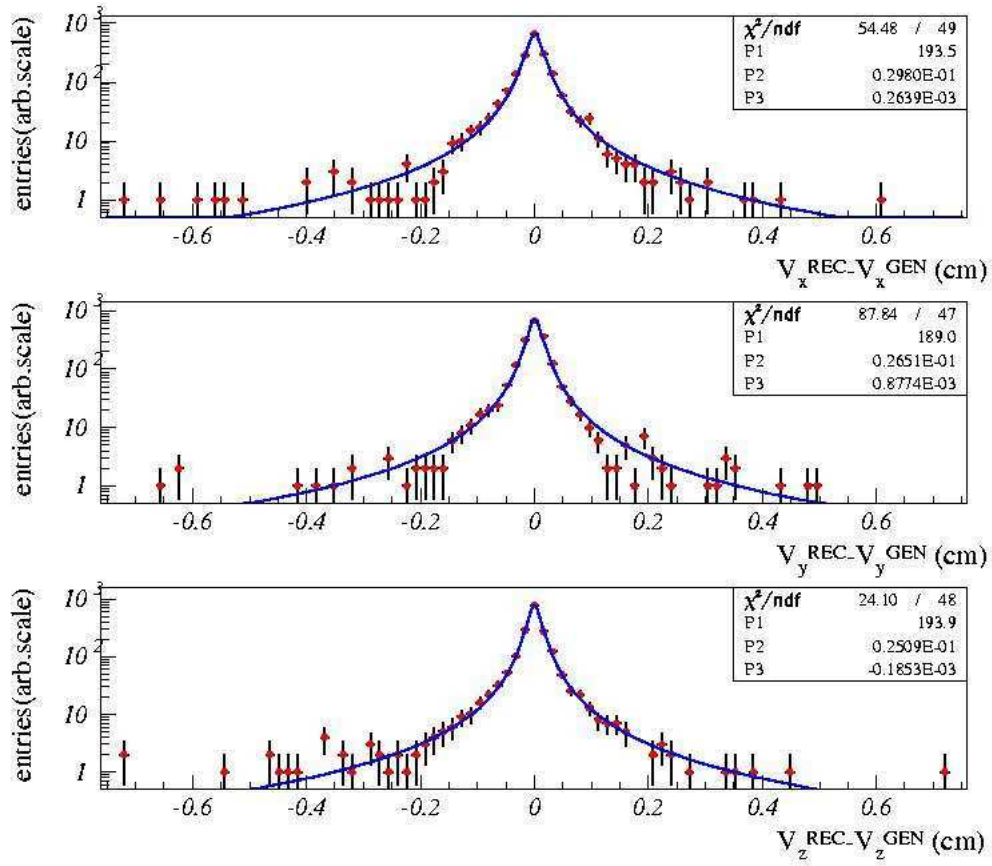


Figure 7.5: Resolution of the secondary vertex coordinates in a $b\bar{b}$ direct photo-production MC sample for 2004 data. The distributions have been parametrized with a Breit-Wigner distribution.

Chapter 8

Event Selection and Secondary Vertex b-Tagging

This chapter is divided into two parts. The first part deals with the dataset and MC samples that have been used, including a description of the selection criteria applied to the events, while the second focuses on the secondary vertex b-tagging algorithm which is the main tool developed within this thesis for the identification of beauty events.

8.1 Event selection

For this analysis of inclusive beauty decays in the photoproduction regime, a dijet event selection is applied to data and MC samples. This selection requires the presence of at least two high- p_T jets in the reconstructed final state. No lepton is required in order to account also for the hadronic b decays keeping the analysis inclusive.

This part of the chapter is organized as follows: the used data and MC samples are listed in subsection 8.1.1. Then the event selection itself is described: first a dijet preselection is done online using a trigger of the third level in the ZEUS trigger chain, as described in subsection 8.1.2. The rest of the selection takes place offline. The photoproduction regime is defined in subsection 8.1.3. Further cuts are applied on variables of interest with the aim to obtain a clean dijet sample with an enriched $b\bar{b}$ component. Section 8.3 summarizes the complete set of cuts applied in the analysis. Finally, control distributions to ensure a good description of the data by the MC simulation are presented in section 8.3.1.

8.1.1 Data and MC samples

The event selection for the first part of the results described in this thesis is based on a data set taken by the ZEUS detector during the running period 2004 with a total integrated luminosity of 38.7 pb^{-1} . During this period, protons were brought into collisions with positrons at a center-of-mass energy of 318 GeV . A first preselection based on the quality of the main detector is done using the ZEUS “take” routines. These routines contain veto flags which are set when a certain detector subcomponent is not working properly during a given run. The veto flags are defined by the component experts in the so called data quality monitoring (DQM) jobs. The main ZEUS data quality flag is called EVTAKES [12] and corresponds to good data quality from the central components CTD, CAL, trigger chain and luminosity measurement. A similar flag based on the quality of the MVD has been applied.

As explained in section 7.3, to determine the decay length with higher precision the beamspot, an average primary vertex position calculated run-by-run, was used instead of the primary vertex. Runs with no calculated beam spot ($\sim 2.23 \text{ pb}^{-1}$) or with an error on the calculation of $\sigma_x > 10 \text{ } \mu\text{m}$ or $\sigma_y > 10 \text{ } \mu\text{m}$ ($\sim 42 \text{ nb}^{-1}$) were removed from the dataset. The final sample used for the analysis has an integrated luminosity of 35.1 pb^{-1} .

The MC samples are needed for the statistical methods applied to obtain the fraction of $b\bar{b}$ in the data as well as for the cross section calculations (Chapter 10). The samples used in this analysis as well as the PYTHIA MC generator used to obtain them are described in detail in chapter 2 and 5. A summary of the generated processes and luminosities is given in table 8.1. The cross section of light flavour photoproduction is much larger than that of beauty photoproduction. This implies that the fraction of beauty events contained in an inclusive MC sample (light and heavy flavour production) is too small. For this reason, beauty MC samples were generated separately and used in addition to the inclusive samples in order to obtain a statistically significant fraction of MC beauty events necessary for this analysis.

Process	$\mathcal{L}[pb^{-1}]$	Process	$\mathcal{L}[pb^{-1}]$
$b\bar{b}$ direct	402.12	inclusive lf. direct	187.60
$b\bar{b}$ resolved	389.49	inclusive lf. resolved	81.19
$b\bar{b}$ excitation in γ	409.67	inclusive excitations contained in the resolved sample	
$b\bar{b}$ excitation in p	440.42		
no cuts		2 jets with $E_T > 4$ GeV and $-3 < \eta < 3$	

Table 8.1: The used MC samples. All samples were obtained using the PYTHIA MC generator. Inclusive lf. samples refers to processes producing both light and heavy flavours.

8.1.2 Trigger

The first step in the selection is to require the presence of at least 2 jets, coming from the hadronization of the heavy quark pair. This selection is carried out by a trigger of the third level of the ZEUS trigger chain (see 4.2.5) and is done online in order to record only physics interactions of interest reducing the amount of recorded background events. Since the trigger system is also simulated in the ZEUS reconstruction program ZEPHYR (**ZEUS PHYsic Reconstruction**), the same trigger selection is applied to the MC samples:

- TLT HFL 5: Inclusive Dijets
 - two jets with $E_T > 4.5$ GeV and $|\eta| < 2.5$
 - CAL: $p_z/E < 0.95$ and $E - Pz < 100$

8.1.3 Photoproduction Selection

The photoproduction regime is defined for photon virtualities $Q^2 < 1$ GeV². In order to select a PHP event sample, DIS events have to be rejected. Within the PHP regime, due to the low Q^2 , the incoming lepton is scattered at a very small angle escaping undetected along the beampipe. On the contrary, in DIS events the lepton is scattered at some measurable angle and can be identified in the detector. This characteristic signature will be exploited to select only photoproduction events by rejecting events with a reconstructed electron in the final state.

To reconstruct electrons, the SINISTRA electron finder is used [1], an algorithm which analyzes the energy deposits in the electromagnetic and hadronic part of the calorimeter and distinguishes between hadronic and electromagnetic clusters. The algorithm consists of two main steps: first, a neural

network provides the probability of a cluster to be hadronic or electromagnetic given the energy of the calorimeter cells. Second, the scattered lepton is selected out of the list of candidates (electromagnetic clusters). Thus, the output of SINISTRA is the lepton candidate having the highest probability P to be the scattered lepton.

SINISTRA can be used in different configurations. For this analysis, the best configuration is the one in which the lepton candidates are reconstructed by using only the CAL information and is not requested to be matched with a track in CTD. This choice is motivated by the fact that the DIS cross section decreases rapidly with increasing Q^2 , so that most of the DIS background to photoproduction is characterized by an outgoing lepton scattered at very low angle. Such a lepton can only be measured by the CAL since it escapes the angular coverage region of the CTD.

With the probability P given by SINISTRA, the energy of the scattered lepton candidate determined by the CAL measurement E'_e and the inelasticity y_{el} calculated following the *electron method* as described in section 6.4, the event is rejected if the candidate with the highest probability fulfills:

- $P > 0.9$,
- $E'_e > 5 \text{ GeV}$ and
- $y_{el} < 0.9$

The last two cuts are done to prevent the rejection of photoproduction events wrongly identified as DIS. Final state pions, electrons or photons, present in a PHP event, can be misidentified as the scattered lepton. Since these particles have lower energies than the scattered lepton, the reconstructed y_{el} which is not the event inelasticity, is larger than for DIS events. Therefore, events with a considerable energy deposit in the electromagnetic CAL but high y_{el} are not excluded from the sample.

Furthermore, a cut on the inelasticity calculated with the *Jacquet-Blonde method* y_{JB} is applied in order to remove residual DIS events. In the definition of y_{JB} (eq. 6.20), the term $\sum_f (E_f - p_{zf})$ sums up the energy E_f and momentum parallel to the beam line p_{zf} for all the final state particles f . In DIS events, the electron energy contributes to this sum even if the electron has not been identified, yielding $y_{JB} \sim 1$ and thus motivating the following cut:

$$0.2 < y_{JB} < 0.8$$

The lower limit is introduced to remove beam-gas interactions characterized by low $(E - P_z)$.

8.2 Secondary Vertex b-tagging

There are many methods to enhance the heavy flavour content of data samples. They not only depend on the characteristics of the detector but also on the physical properties of the event being studied. One of these methods is called secondary vertex tagging. The analysis reported in this thesis studies beauty production exploring the feasibility and application of this technique for beauty identification in ZEUS.

The secondary vertex tagging exploits the relatively long lifetime of b hadrons to reconstruct a secondary vertex significantly displaced from the beam spot or primary vertex. In fact, the transverse decay length of a particle of mass m , mean lifetime τ and transverse momentum p_T is approximately given in the laboratory frame by

$$L_{xy} = \frac{p_T}{m} c\tau. \quad (8.1)$$

B hadrons (masses between $5279 - 5624 \text{ MeV}$, $c\tau$ between $368 - 501 \mu\text{m}$ [34]), with a momentum of a few GeV travel a short distance before decaying. Their decay products originate from the decay vertices (commonly called secondary vertices), and not from the primary vertex like most produced particles in the event. Thus, secondary vertices, reconstructed from displaced tracks fulfilling certain requirements, can be discriminated from the primary vertex. Figure 8.1 illustrates the reconstruction of a secondary vertex inside a jet. The main background to be taken into account for this tagging method originates from hadrons containing a c -quark since in the detector they can have decay lengths similar to those of the B -hadrons.

The choice to use a method based on the reconstruction of the secondary vertex instead of the impact parameter of single tracks was taken for several reasons. First, to exploit the full three dimensional information provided by the MVD. In fact, while a reconstructed secondary vertex is a three dimensional object, the impact parameter is usually defined in two dimensions. Second, to reconstruct a decay vertex is to reconstruct a B hadron which implicitly means that all decay products of the B hadron are detected so that this method can be better used in inclusive approaches where no lepton is required in the final state, as it is the case in semileptonic analyses. And third, while the impact parameter method has more statistical power and

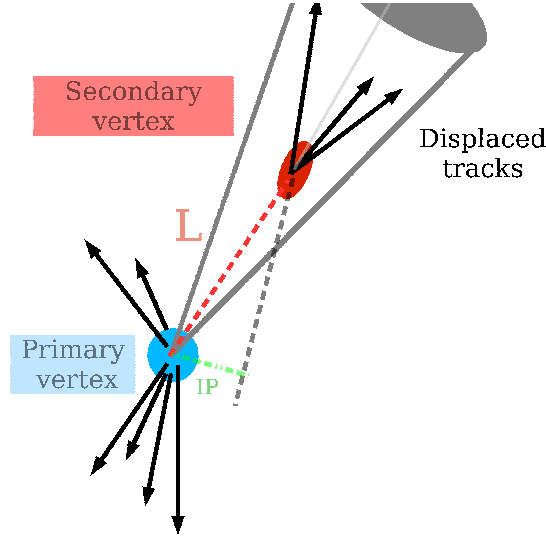


Figure 8.1: Schematic representation showing how the decay vertices (secondary vertices) of long-lived particles can be reconstructed from the displaced tracks w.r.t. the primary vertex. The observable to be used in this analysis is thus the decay length L defined as the distance between the primary and the secondary vertex. Specifically the transverse decay length L_{xy} will be used (see text). Also the impact parameter (IP, in other analysis is denoted as δ), defined as the distance of closest approach of the particle track to the primary vertex, is depicted in the figure for one of the tracks.

is conceptually and technically simpler, it is characterised by larger backgrounds than the methods based on the secondary vertices.

Hereafter the main steps of the secondary vertex algorithm will be discussed. Since the reconstruction of the primary vertex was already discussed in detail in Chapter 7, it will be skipped here. Intermediate steps such as the track and jet selection will be reviewed in sections 8.2.1 and 8.2.2. Section 8.2.3 describes the track-jet association as well as the determination of the secondary vertex position. Once a secondary vertex is found, the two dimensional decay length L_{xy} is calculated. Section 8.2.4 describes how it is obtained and used to distinguish b quarks from c and light quarks.

8.2.1 Track selection

For every event, only high energy “ZTT” tracks (see Sec. 6.1) are selected. In order to reduce the amount of tracks which are not well measured, also a number of track quality cuts are applied. Table 8.2 summarizes the track selection cuts. All the cuts are designed to enrich the signal content of

the sample. The transverse momentum requirement, $p_T > 0.5 \text{ GeV}/c$, is motivated by the fact that decay products of a B -hadron are expected to have bigger transverse momentum values than that of the background as well as by the vertex resolution which improves with the momentum of the tracks. The other cuts are standard requirements for a good quality reconstructed track.

Track type	ZTT
Transverse momentum	$p_T > 0.5 \text{ GeV}/c$
Number of superlayers in CTD	> 3
Total number of hits inside MVD	≥ 4
χ^2 of the track	$\chi_{track}^2 < 100$

Table 8.2: Selection criteria for tracks used in the b-tagging algorithm.

8.2.2 Jet selection

Motivated by the kinematic characteristics of the B-hadron decays, the events are selected requiring at least 2 reconstructed jets (see Sec. 6.3) in the final state with transverse momenta $P_{T1}^{jet} > 7 \text{ GeV}$ and $P_{T2}^{jet} > 6 \text{ GeV}$ and pseudorapidity $|\eta^{jet}| < 2.5$. The η range follows from the geometrical structure of the detector. Table 8.3 summarizes the jet selection cuts.

Jet algorithm	Kt-Algorithm E recombination scheme
Transverse momentum Jet 1	$p_{T1} > 7 \text{ GeV}/c$
Transverse momentum Jet 2	$p_{T2} > 6 \text{ GeV}/c$
Pseudorapidity	$ \eta_{1,2}^{jet} < 2.5$

Table 8.3: Selection criteria for jets used in the b-tagging algorithm.

8.2.3 Track-Jet Association and determination of the secondary vertex position

Selected tracks are associated to jets using the variable ΔR , defined as the distance between the two objects in the (η, ϕ) plane (Eq. 8.2). By requiring $\Delta R < 1$ for every combination track-jet, the track is associated to the closest jet in this space (see Figure 8.2).

$$\Delta R = \sqrt{(\eta_{trk} - \eta_{jet1,2})^2 + (\phi_{trk} - \phi_{jet1,2})^2}. \quad (8.2)$$

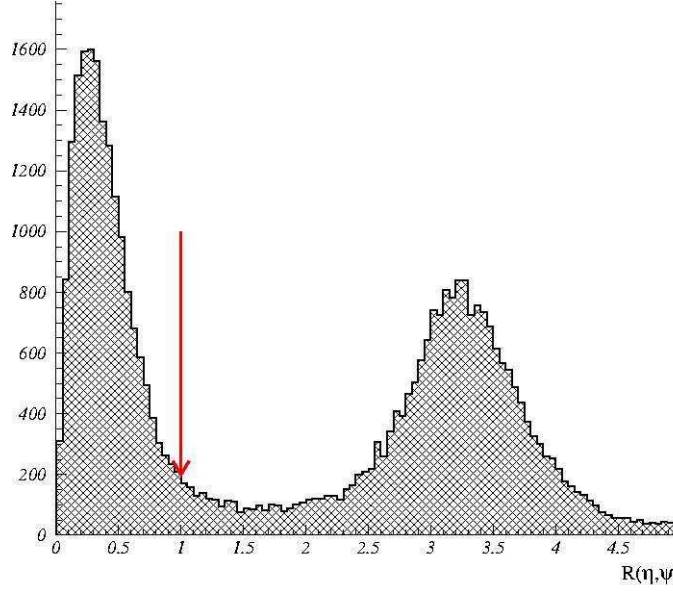


Figure 8.2: Distance ΔR between track and jet. $\Delta R < 1$ is required to define the track-jet association.

For every jet with at least two selected tracks, an attempt to reconstruct a secondary vertex is done. Secondary vertices are reconstructed using the user **VerteX Utility** (VXU) algorithm [43] described in detail in chapter 7. If the algorithm finds a vertex, the fit procedure outputs the three vertex coordinates and the covariance matrix. A final cut is applied in order to reject secondary interactions and photon conversions in the beampipe or in the MVD material. This cut consists in removing all vertices which lie at more than 1 cm in the $x - y$ plane and 30 cm in the z plane from the nominal interaction point (see 7.4).

8.2.4 Decay length distributions

Once a secondary vertex is found, the decay length L is calculated as the distance between the beam spot (average primary vertex, section 7.3) and the secondary vertex. We recall that the beam spot was chosen over the primary vertex due to the higher precision on its position. However, the beam spot is only calculated in the transverse $x - y$ plane which leads to the use of the decay length in the transverse plane L_{xy} instead of the 3-dimensional L . The transverse decay length is defined as (see Fig. 8.3):

$$L_{xy} = |\vec{S} - \vec{P}| \text{sign}((\vec{S} - \vec{P}) \cdot \hat{j}) \quad (8.3)$$

where \vec{P} and \vec{S} denote the position vectors of the primary¹ and secondary vertex and \hat{j} the vector along the jet axis, all of them in the transverse plane XY . The decay length defined above is given a specific sign which is determined from the relative position of the secondary vertex with respect to the primary vertex and the jet axis. If the secondary vertex is in the same hemisphere² as the primary vertex and the jet, L_{xy} is positive, otherwise (the vertex is in the opposite hemisphere of the detector i.e. behind the jet) it is defined to be negative (see Fig. 8.3). Such vertices are not consistent with heavy flavour decays and are due to the finite resolution of MVD which sometimes gives a wrong reconstruction of the secondary vertex behind the primary vertex.

As expected from the lifetime difference between charm hadrons and beauty hadrons, the shape of the decay length distributions is significantly different for b jets than for charm and light flavour jets. This can be observed in fig. 8.4 where the decay length was calculated for different MC samples: beauty, charm and light flavour respectively. Tracks in a jet coming from a light quark should originate from the primary vertex. Due to the finite vertex fitting resolution, these vertices are reconstructed as decay vertices very close to the primary interaction yielding a transverse decay length that has equal probability to be either positive or negative signed. Thus, while the decay length distribution for light flavour is almost symmetric, it becomes slightly asymmetric towards positive values for the charm sample and completely asymmetric for the beauty jets. This discriminative power of L_{xy} is shown again in Fig. 8.5(a). In figure 8.5 (b) the reconstructed L_{xy} spectra for a beauty MC sample appears superimposed to the true L_{xy} calculated for the same sample, exposing the point stated before that the negative values of L_{xy} correspond to a wrong reconstruction given by the tracking detectors. The fraction of events with negative L_{xy} gives a indication of the number of fake secondary vertices yielding a positive L_{xy} .

It should be noted that a more sophisticated analysis would imply to use the decay length significance rather than the decay length itself. The significance is defined as $L_{xy}/\sigma_{L_{xy}}$ where $\sigma_{L_{xy}}$ denotes the error on L_{xy} and is computed from the covariant matrix elements of the primary and secondary vertex positions. However, this analysis was carried out with an early version of the reconstruction of the 2004 data which had strong limitations [70, 90] due to misalignments and calibration problems, not enough accurate understanding of the tracking/vertexing procedures and wrong treatment of

¹From now on, and so long as nothing else is specified, primary vertex denotes the average primary vertex i.e. the beam spot.

²The jet axis is used to divide the event into two hemispheres delimited by a plane perpendicular to the jet axis at the position of the primary vertex.

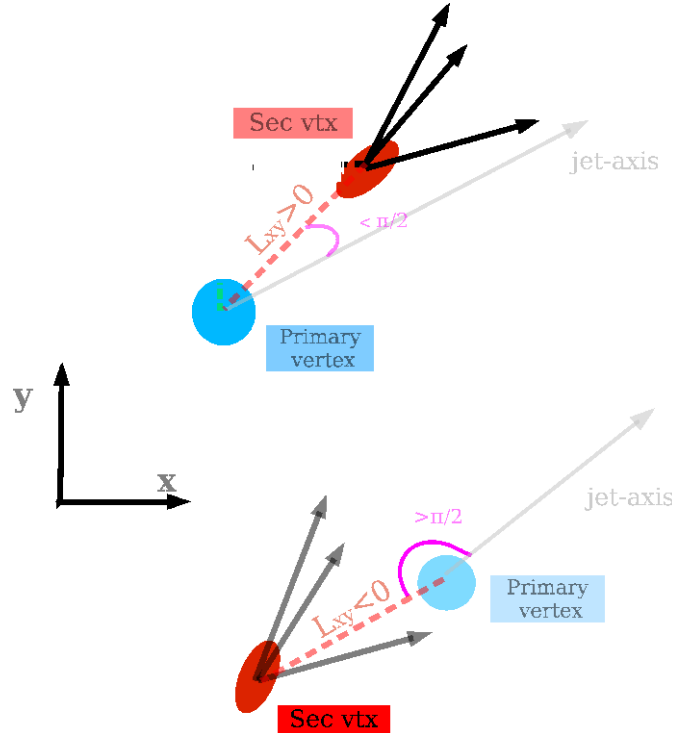


Figure 8.3: Schematic representation of the transverse decay length L_{xy} . L_{xy} is given a sign which is determined from the relative position of the secondary vertex with respect to the primary vertex and the jet direction. All objects in the figure are projected on the xy -plane. If the secondary vertex is in the same hemisphere than the primary vertex and the jet, the sign is positive (above), otherwise it is negative (below)

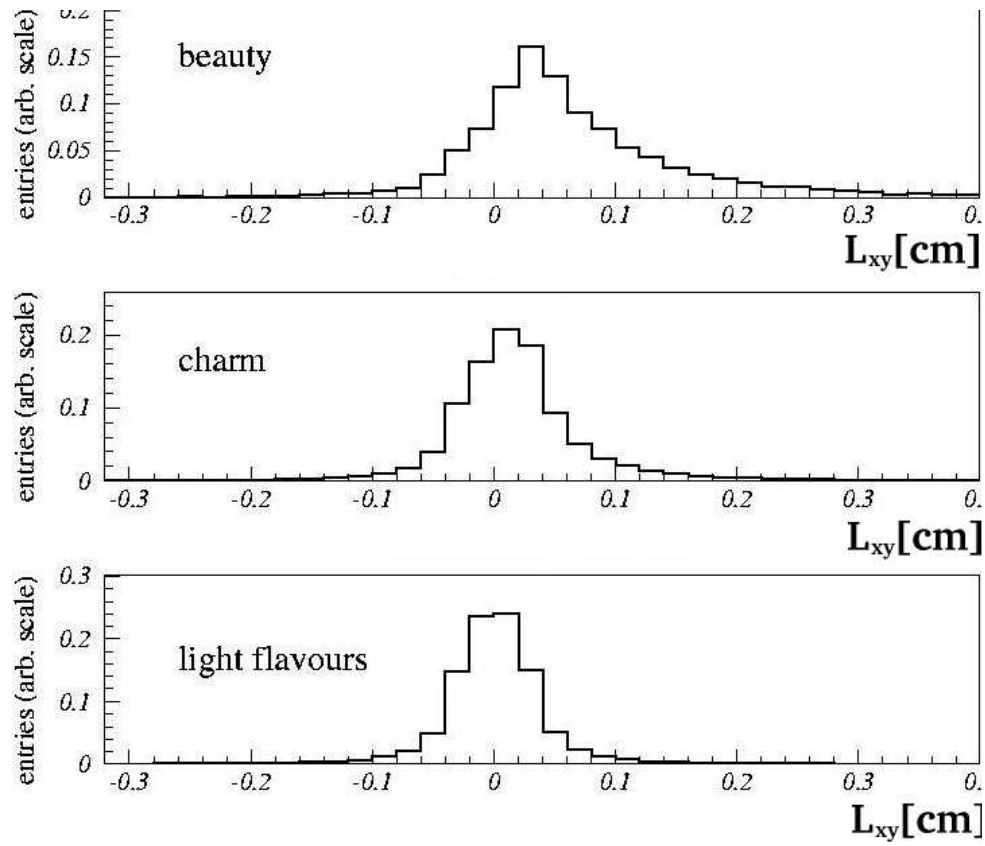


Figure 8.4: Decay length distributions for beauty, charm and light flavour samples. The asymmetry of L_{xy} toward positive values for the beauty sample in contrast to the other samples emphasizes the discriminative power of this variable.

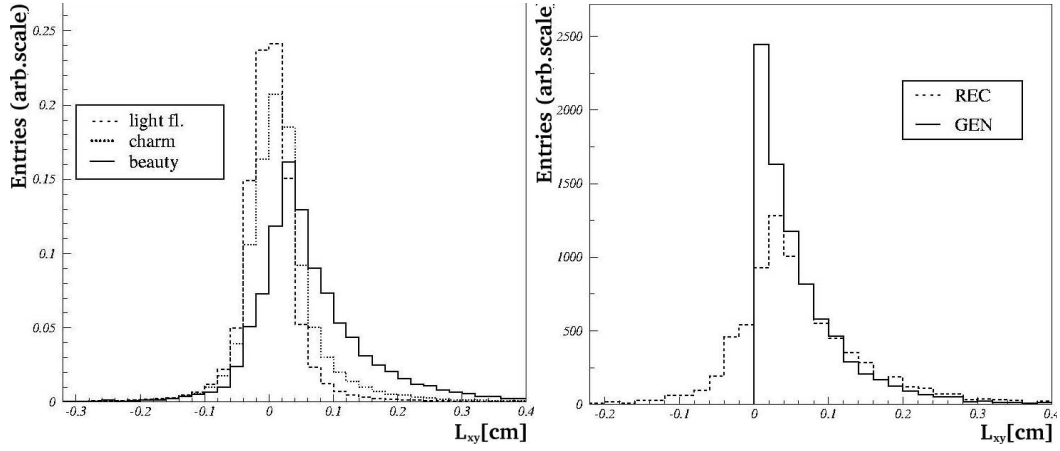


Figure 8.5: Superimposed decay length distributions of (left) beauty, charm and light flavour MC reconstructed samples, (right) reconstructed and true beauty decay length

the errors that makes the determination of $\sigma_{L_{xy}}$ not reliable. Moreover, it will be seen in the next chapter that the decay length distributions show an important disagreement between data and Monte Carlo making requiring a smearing of the MC decay length distribution. First results from an improved analysis using the better reconstructed 2005 data will be shown in chapter 12.

8.3 Final Selection

Summarizing the complete selection, the following cuts have been applied to the data and MC samples used for the first analysis reported in this thesis:

- **Trigger preselection:** inclusive dijets
 - two jets with $E_T > 4.5 \text{ GeV}$ and $|\eta| < 2.5$
 - CAL: $p_z/E < 0.95$ and $E - Pz < 100$
- **Photoproduction selection:**
 - reject event if an electron is reconstructed with:
 - * $Prob(e) > 0.9$,
 - * $E'_e > 5 \text{ GeV}$,
 - * $y_e < 0.9$.

- cut on inelasticity using the Jacquet-Blondel method $0.2 < y_{JB} < 0.8$
- cut on the square momentum transfer $Q^2 < 1$.
- **Jet selection:** at least two jets reconstructed with the KTCLUS algorithm with
 - $p_T^{jet1} > 7 \text{ GeV}$ and $p_T^{jet2} > 6 \text{ GeV}$,
 - $|\eta^{jet12}| < 2.5$.
- **Track selection:** only “ZTT” tracks were selected fulfilling the following requirements
 - $p_T^{trk} > 0.5 \text{ GeV}/c$,
 - number of SL in CTD > 3 ,
 - total number of hits inside MVD ≥ 4 ,
 - $\chi_{trk}^2 < 100$
- **Track-Jet association:** each track is associated to the closest jet requiring
 - $\Delta R_{trk-jet} < 1$
- **Additional cuts:** the following cuts were applied to reject secondary interactions of no interest
 - $|Z_{primary-vertex}| < 50 \text{ cm}$,
 - Remove all secondary vertices at more than 1 cm in the $x-y$ plane and 30 cm in the z plane from the nominal interaction point,
 - Remove all tracks which lead to a vertex made of two oppositely charged tracks with mass compatible with K_s^0 and Λ

8.3.1 Control distributions

In this section, data distributions are compared to the MC predictions for the most relevant kinematic variables of the tracks (Fig. 8.6) and jets (Fig. 8.7) used in this analysis. All the MC distributions correspond to the inclusive “light flavour” sample (i.e. contains light flavour, charm and beauty) and were normalized to the corresponding data areas. In all cases, MC and data distributions are compared after the event selection explained in the last

section. Also CTD and MVD quantities are checked in Fig. 8.8 since the track selection is directly related to those variables. Reasonable agreement between data and MC simulation is observed in all given variables ensuring the feasibility of the method.

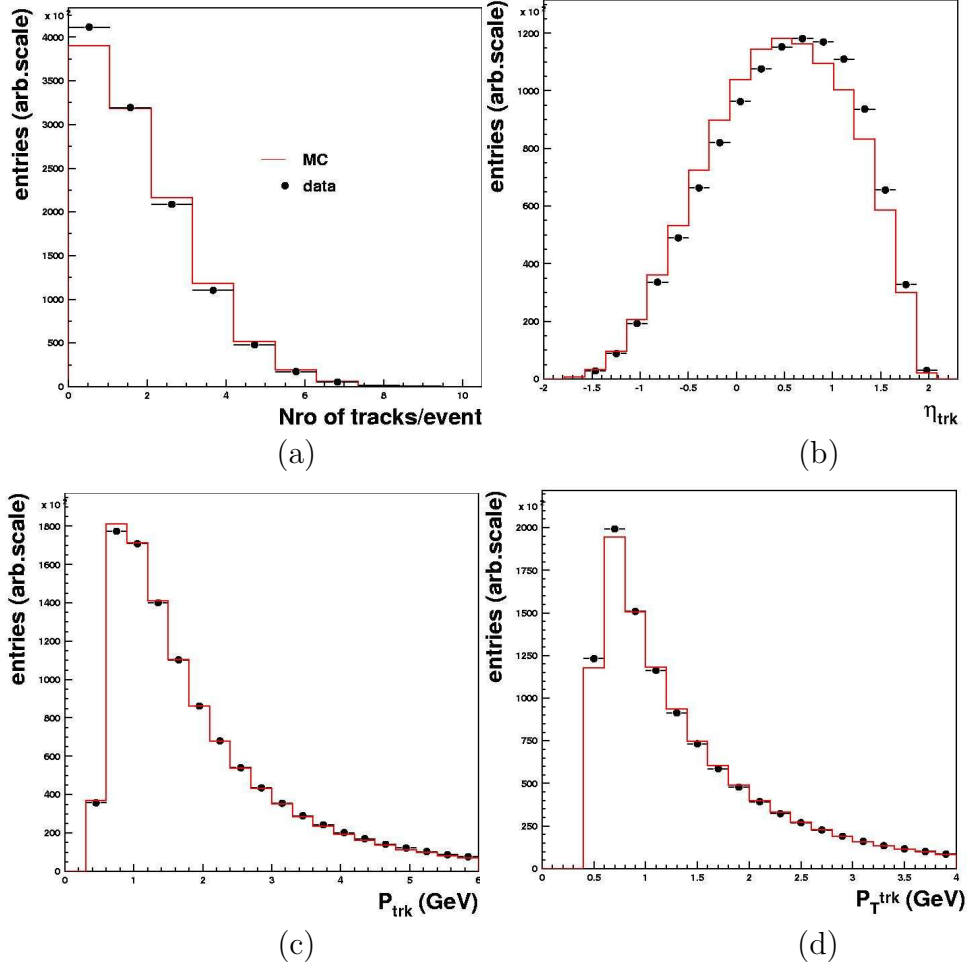


Figure 8.6: Control distributions of the track related variables. Only tracks passing the final selection (see text) are considered here: (a) Number of tracks in the event, (b) pseudorapidity η_{trk} , (c) momentum of the track and (d) transverse momentum p_T of the tracks. All the Monte Carlo histograms are normalized to the data area.

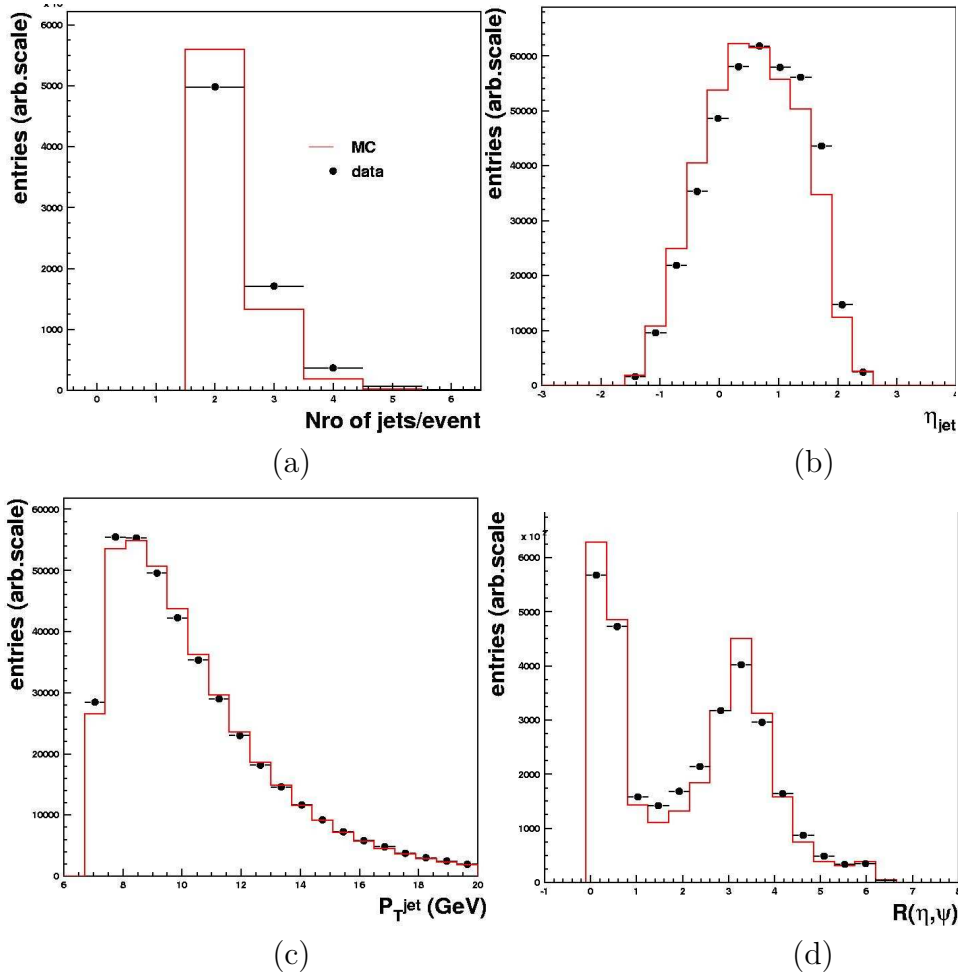


Figure 8.7: Control distributions of the jet related variables: (a) Number of jets in the event, (b) pseudorapidity η_{jet} , (c) transverse momentum p_T of the jets and (e) the distance ΔR between track and jet. All the Monte Carlo histograms are normalized to the data area.

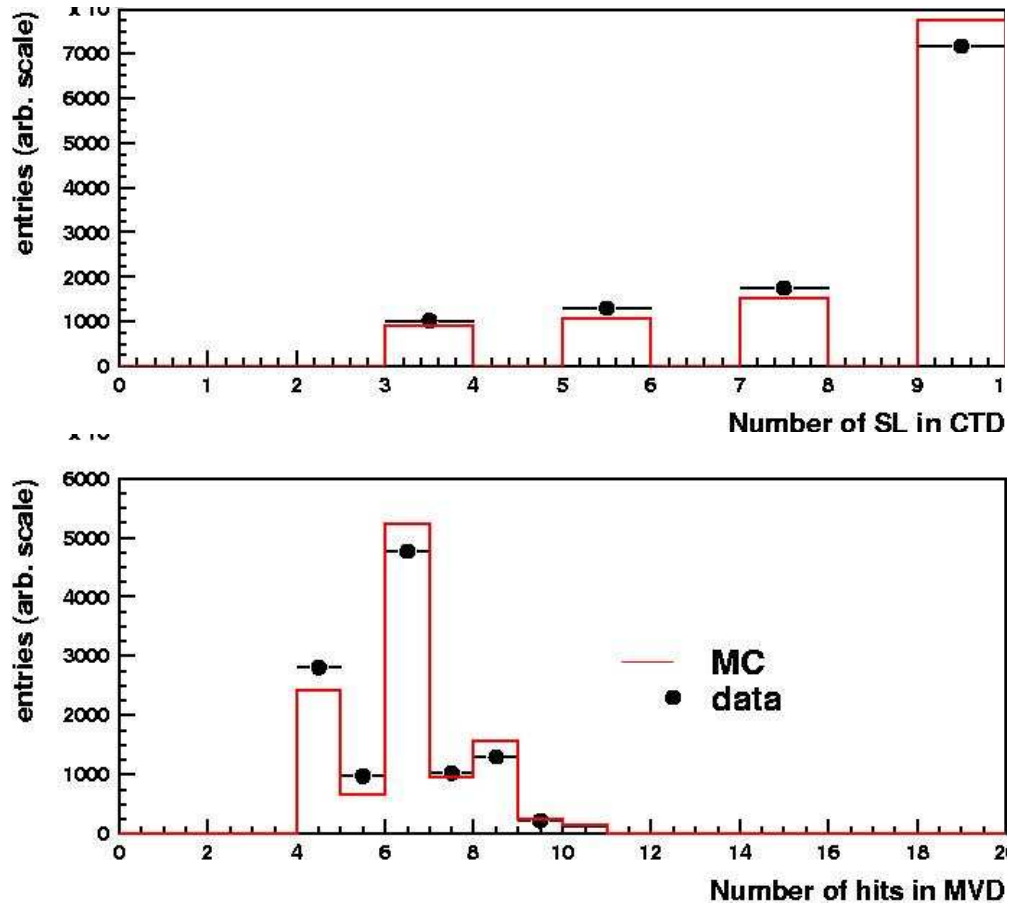


Figure 8.8: Control distributions of the track variables related to the MVD-CTD:(a) Number of Super Layers in the CTD crossed by the selected tracks, (b) Number of MVD hits used in the reconstruction of the track. All the Monte Carlo histograms are normalized to the data area.

Chapter 9

Extraction of Beauty Signal

This chapter describes the method used in this thesis for the determination of the beauty fraction from the selected data sample. The decay length, already discussed in the previous chapter, is the main variable for extraction of the beauty fraction. However, in order to achieve an efficient separation of the b -quarks from other flavours the discriminating power of the decay length will be enhanced by using, in a combined tagging algorithm, another variable sensitive to b -quarks: the secondary vertex mass M_{SV} .

During the development of this analysis, several problems were encountered. Their nature rely on the problems, already known, of the 2004 dataset (see Sec. 8.2.4) as well as a poor description of some of the main variables by the Monte Carlo. Consequently, before the signal could be extracted, different corrections were applied in an attempt to solve these problems. The problems and the corrections will be described in detail along this chapter as well.

9.1 Analysis strategy

The analysis presented in the first part of this thesis is based on the decay length and the secondary vertex mass, variables which, in turn, are built up from reconstructed secondary(decay) vertices. Unfortunately, the track and vertex reconstruction algorithms were not yet completely understood when this study was performed. This added to the known problems of the 2004 dataset, makes a succesfully reconstruction of decay vertices rather difficult. In fact, the contribution of misidentified decay vertices is higher than expected. The immediate consequence of this wrong vertex reconstruction is that the shapes of the decay length L_{xy} and the secondary vertex mass M_{SV} distributions respectively between beauty and charm are not sufficiently different and therefore, a good separation of beauty from charm and light

flavours events becomes impossible by an individual use of any of these two discriminating variables. This fact led to the idea of using both variables together instead of separately. Indeed, combining both variables, L_{xy} and M_{SV} , provides results which are not only satisfactory within the framework of this analysis, constrained by its limitations as already discussed along this thesis, but also promising for future analysis. The discriminative capacity of a combined use of both variables is described in detail in section 9.3.

Furthermore, it was also seen in the previous chapter that the transverse decay length L_{xy} is given a sign which depends on the relative position of the secondary vertex with respect to the primary vertex and the jet axis. This sign is used to distinguish measured decay lengths resulting from real secondary vertices from those which are associated with a wrong reconstruction due to the resolution effects of the detector. Thus, subtracting the negative side of the decay length distribution L_{xy}^- from the positive side L_{xy}^+ allows to cancel the systematic errors associated to the tracking detector resolutions. Moreover, this subtraction eliminates (except fluctuations) the light flavour contribution so that the $(L_{xy}^+ - L_{xy}^-)$ distribution has now only 2 parameters, corresponding to beauty and charm, to be determined using a statistical method, as it will be explained in detail in section 9.5, based on a least squares simultaneous fit. Each MC flavour contribution is then scaled by the obtained factors which give the best fit to the measured $(L_{xy}^+ - L_{xy}^-)$ distribution. Note that along this chapter an indiscriminate use of L and L_{xy} in the figures is done, referring always to the transverse decay length.

9.2 Monte Carlo smearing of the L_{xy} distribution

The reliability of the secondary vertex algorithm depends on the Monte Carlo description of all the variables used directly or indirectly in the analysis. At the end of the last chapter, control distributions were shown for the most important track and jet variables that are used in the calculation of L_{xy} and a good agreement is found in all of them. However, a comparison of the decay length distribution between the data and MC simulations (Fig. 9.3 (a)) shows substantial discrepancies, the data distribution being significantly wider than the MC distribution. Reasons for this disagreement could be:

- the dead material is not simulated with enough accuracy in the Monte Carlo,
- the intrinsic hit and track resolutions are not described correctly in the

Monte Carlo simulation,

- and the alignment of the Micro Vertex Detector needs to be improved.

The improved analysis described in chapter 12 suggests that these are some of the reasons responsible for the disagreements.

Correspondingly, the simulated decay length was smeared in order to achieve a satisfactory agreement between data and MC distributions. The smearing was done according to studies performed in [77, 76]. A brief description of the smearing procedure is given below. In order to determine the best smearing

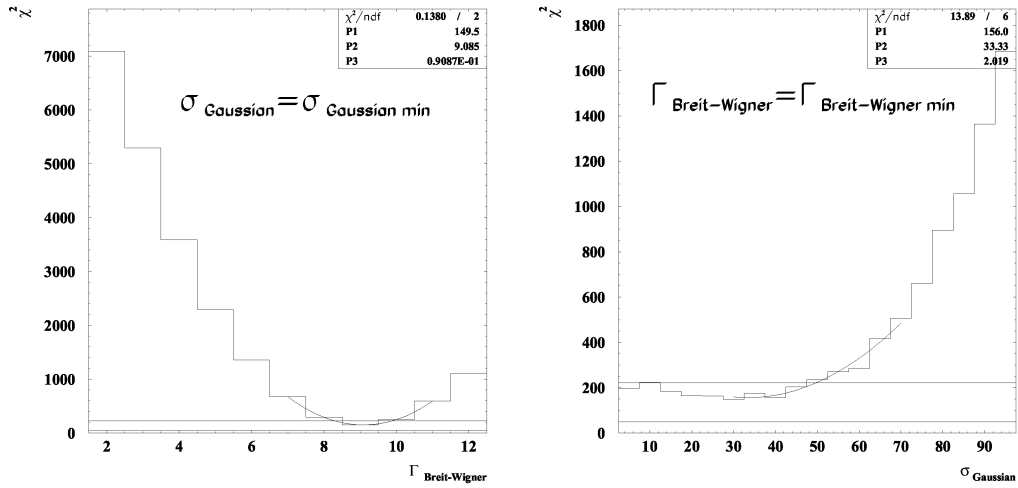


Figure 9.1: Example of a 2D plot of the χ^2 distribution for $0.5 < p_T < 1 \text{ GeV}$. On the right the distribution of χ^2 is plotted as a function of the Gaussian σ , keeping the value of the Breit-Wigner Γ constant at its minimum. On the left the distribution of the χ^2 is plotted as a function of the Breit-Wigner Γ , keeping the value of σ constant at its minimum [77, 76].

to be applied to the MC, the negative side of the data distribution was fitted with different functional forms. A double convolution of the MC distribution with a Gaussian and a Breit-Wigner was found to be the best fit to the data. It can be expressed as:

$$F^1(x) = \int dz F(z) B(x - z) \quad (9.1)$$

$$F^2(x) = \int dy \int dz F(z) B(y - z) G(x - y) \quad (9.2)$$

where B indicates the Breit-Wigner function and G the Gaussian function.

The exact values for the widths of the two distributions were obtained testing different combinations of the σ of the Gaussian and Γ of the Breit-Wigner until the minima were found. For this purpose, 2D distributions of the χ^2 of the fit were done in which the χ^2 of the fit was plotted as a function of the Gaussian width keeping the gamma of the Breit-Wigner constant at its minimum. The Gaussian width was determined by fitting the local minimum area with a parabolic fit. The same procedure was done to obtain the Breit-Wigner gamma. The whole study was performed in different P_t bins. An example of these distributions for the first bin of p_T can be seen in fig. 9.1.

To determine the final minima of σ and Γ for the whole kinematic range, the distributions of the above minima as a function of P_T were fitted using a combination of an exponential and a constant function in the case of the Breit-Wigner and with a constant term for the Gaussian (see Fig. 9.2):

$$\Gamma_{BW}(P_T) = \exp(a + b \cdot P_T) + c \quad (9.3)$$

$$\sigma_{Gauss}(P_T) = d \quad (9.4)$$

where the parameters are $a = 3.0275$, $b = -2.2336$, $c = 4.8768$ and $d = 0.0048781$. These values are used to correct the impact parameter measurements as follows:

$$IP_{measured} = IP_{original} + sme_{BW} + sme_{Gauss} \quad (9.5)$$

where sme_{BW} and sme_{Gauss} are the empiric smearing values coming from the Breit-Wigner and Gaussian function respectively:

$$sme_{BW} = N_{RAN}^{BW}(N_{RAN}^{UNI}) \times \Gamma_{BW}(P_T) \times C. \quad (9.6)$$

and

$$sme_{Gauss} = \sigma_{Gauss}(P_T) \times N_{RAN}^{Gauss} \quad (9.7)$$

In Eq.9.6 N_{RAN}^{BW} denotes a random number generated with Γ_{BW} equals to 1, N_{RAN}^{UNI} is an uniformly generated random number from 0 to 1 and C is a constant. While in Eq.9.7 N_{RAN}^{Gauss} is a random number generated according to a Gaussian distribution with σ equals to 1.

These studies are based on the impact parameter(or distance of closest approach) rather than the decay length so the smearing is not optimized for this analysis. Nevertheless, the application of this smearing implied a remarkable improvement in the agreement between data and MC. Figure 9.3

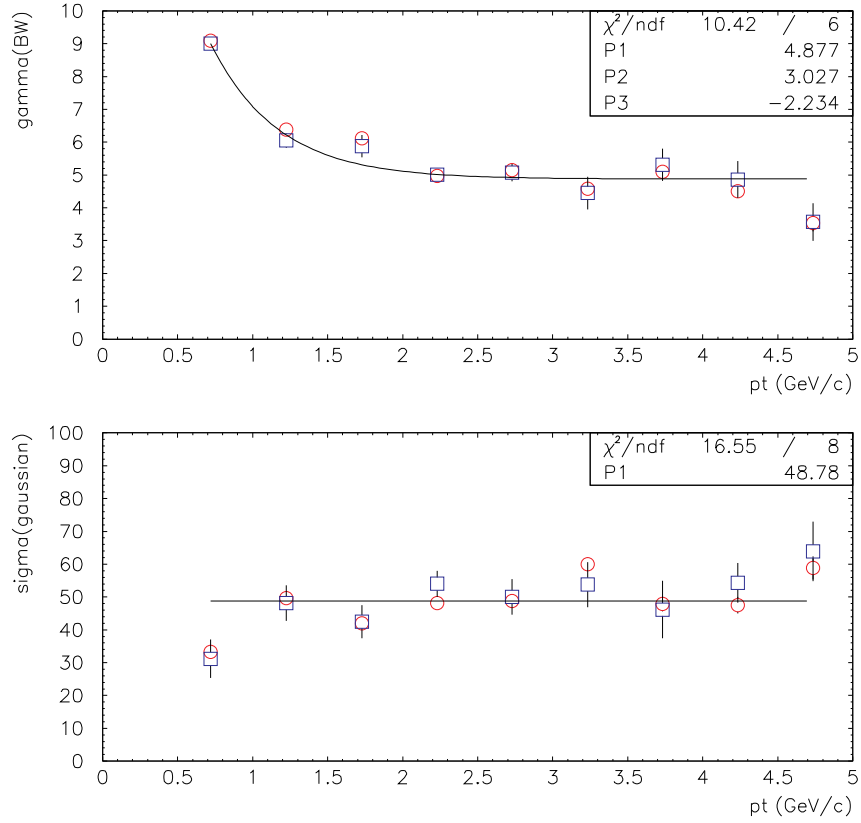


Figure 9.2: Distribution of the Breit-Wigner gamma (top) and Gaussian sigma (bottom) in bins of p_T . The distributions were fitted with a combination of an exponential and a constant function (top) and with a constant function (bottom). The minima are estimated using a likelihood (circles) and bayesian (squares) approach (see [77] for more details).

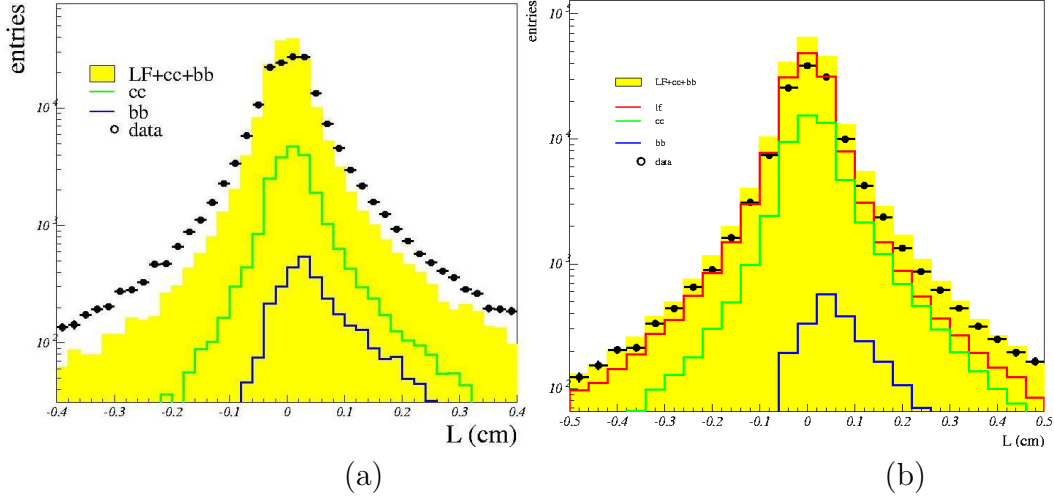


Figure 9.3: Comparison of the decay length distribution for the MC and data before (a) and after the smearing has been applied (b). The green and the blue line represents the contribution coming from charm and beauty events respectively. Note that the binning is different in both histograms.

(b) shows the decay length distributions after the smearing has been applied. The improvement is clearly seen. The apparent normalization problem seen in this figure is solved as it will be explained in the next section.

9.3 The secondary vertex mass M_{SV}

The variable M_{SV} is defined as the total invariant mass of the tracks associated with the secondary vertex. It is calculated by assuming that each track has the mass of a pion and it is sensitive to the quark flavour. The invariant mass distribution at the secondary vertex corresponding to a jet generated by a c -quark falls rapidly above the D-meson mass, which is about $1.8 \text{ GeV}/c^2$, while the mass in a b -jet can go up to $5 \text{ GeV}/c^2$ (B -hadron mass). Figure 9.4 (a) shows the invariant mass distribution for a MC sample (the different flavour contributions can be seen) compared to the data. The values for the invariant mass of c -jets that are above the limit $1.8 \text{ GeV}/c^2$ are due to tracks incorrectly attached to the secondary vertex (misidentified decay vertices). As it can be seen from this figure, the MC does not describe properly the data M_{SV} distribution. To correct for this discrepancy between MC and data, the $[data/MC]$ ratio is calculated from the M_{SV} distributions and is applied as a reweight factor to all the MC distributions. Figures 9.4 (b) and (c) show the M_{SV} and L_{xy} distributions respectively after the reweighting has been

applied. The normalization problem observed in the L_{xy} distribution (see Fig. 9.3 (b)) is also solved after applying this correction.

In figure 9.4 (d), the already mentioned discriminative power of the combined use of both variables is illustrated. The M_{SV} spectrum is plotted for all vertices with a transverse decay length L_{xy} bigger than 0.1 cm. As it can be seen, the signal (beauty) to background (light flavour and charm) ratio improves considerably after this cut and a beauty enriched sample can be obtained.

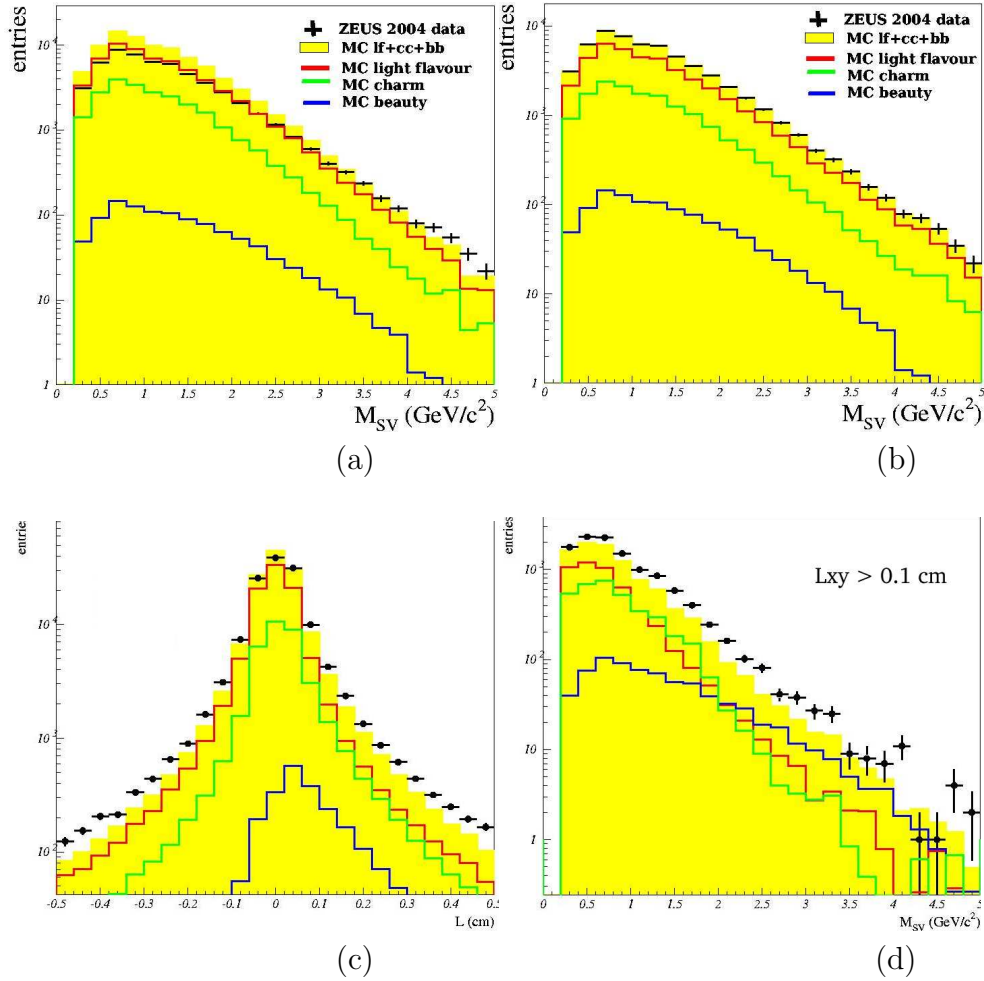


Figure 9.4: Secondary vertex mass distribution of the data compared to the different MC flavours contributions, before (a) and after the Monte Carlo is reweighted to the data distribution (b) (see text). Transverse decay length distribution after the reweighting (c). Secondary vertex mass distribution after applying a cut on $L_{xy} > 0.1 \text{ cm}$; a beauty enriched sample is obtained at large masses.

9.4 The $(L_{xy}^+ - L_{xy}^-)$ distribution

When the strategy of this analysis was discussed at the beginning of this chapter, it was seen that building up the variable $(L_{xy}^+ - L_{xy}^-)$ has many advantages since it cancels some systematic errors and also eliminates the light flavour contribution making a simultaneous 2-parameter fit of two distributions possible. This is the keystone for the extraction of the beauty fraction. The two distributions to be fitted simultaneously are the $(L_{xy}^+ - L_{xy}^-)$ spectra in two different mass regions. This proceeding is motivated by reasons analogous to the ones given in section 9.1 for the decay length L_{xy} : the $(L_{xy}^+ - L_{xy}^-)$ distribution alone can not be used to determine a meaningful beauty fraction since the MC distribution shapes for charm and beauty are too similar but it can be used in combination with the secondary vertex mass. For this purpose, the $(L_{xy}^+ - L_{xy}^-)$ variable was divided in the three bins of M_{SV} : $0.7 \leq M_{SV} < 1.4 \text{ GeV}/c^2$, $1.4 \leq M_{SV} < 2. \text{ GeV}/c^2$ and $M_{SV} \geq 2. \text{ GeV}/c^2$. In each histogram from figure 9.5 the $(L_{xy}^+ - L_{xy}^-)$ distribution is plotted for each bin of M_{SV} . The size of the bins has been chosen according to the mass of the jets containing a c or b quark. Thus, the second mass bin is expected to be dominated by charm while in the third bin the main contribution comes from beauty events. The $(L_{xy}^+ - L_{xy}^-)$ distribution for $M_{SV} \leq 0.7$ is not considered in the analysis since the MC simulation does not describe the data correctly in this region.

The $(L_{xy}^+ - L_{xy}^-)$ correction for charm

The statistical method applied to extract the beauty fraction in the sample (explained in next section) relies on the MC description of the variable to be fitted: $(L_{xy}^+ - L_{xy}^-)$. However, in figure 9.5 a big disagreement between MC and data for the low region of $(L_{xy}^+ - L_{xy}^-)$ is observed. This region is of strong importance since it is dominated by the charm contribution and therefore, needed in the simultaneous 2-parameter fit of 2 distributions to estimate the charm fractions. Hence, a correction is applied. It was chosen to use the $(L_{xy}^+ - L_{xy}^-)$ distribution corresponding to the first bin of M_{SV} to adjust empirically the charm contribution. The correction consists in calculating the factor $[data/MC_{c\bar{c}}]$ in this first M_{SV} bin, where $MC_{c\bar{c}}$ denotes the MC charm distribution, and using it to scale the charm distributions of the other two bins of M_{SV} . The resulting distributions are the ones to be used for the fit. As a consequence, since the charm distribution has been adjusted “by hand”, the charm factor which will be obtained from the fit can not be used anymore to determine the fraction of charm in the sample but it provides the necessary tool to extract the beauty factor.

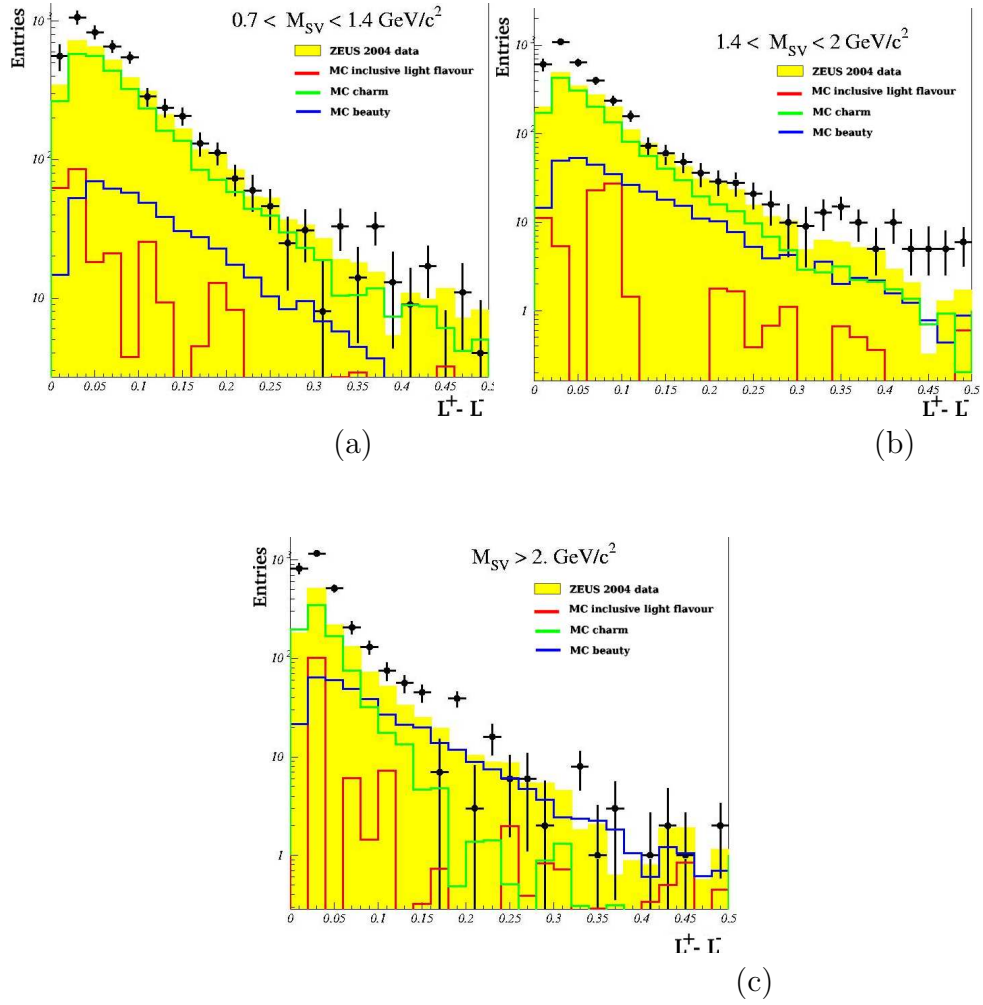


Figure 9.5: $(L_{xy}^+ - L_{xy}^-)$ distributions in bins of M_{SV}

9.5 Extraction of the beauty fraction

The determination of the beauty fraction is performed by the simultaneous fit of the $(L_{xy}^+ - L_{xy}^-)$ beauty and charm MC distributions in the second and third bin of M_{SV} . The fit procedure is done using the MINUIT package [42] which is a tool based on finding the minimum value of a multi-parameter function and analyzing the shape of the function around the minimum. The function to minimize is the binned χ^2 defined as:

$$\chi^2 = \sum_i \frac{(n_i - P_1 N_i^{c\bar{c}} - P_2 N_i^{b\bar{b}})^2}{n_i - (P_1 \sigma_i^{c\bar{c}})^2 - (P_2 \sigma_i^{b\bar{b}})^2} \quad (9.8)$$

where $N_i^{c\bar{c}}$ and $N_i^{b\bar{b}}$ are the contents of the i^{th} bin in the beauty and charm $(L_{xy}^+ - L_{xy}^-)$ distributions and n_i the content of the i^{th} bin of the data. The free parameters P_1 and P_2 can be interpreted as the scaling factor by which the beauty(charm) MC distributions must be multiplied in order to obtain the beauty(charm) fraction in the data sample i.e. each flavour contribution is scaled by the factors which give the best fit to the measured $(L_{xy}^+ - L_{xy}^-)$ distribution. The statistical errors of the data and MC simulation are considered in the fit.

Thus, the minimization is done with respect to the two free parameters P_1 and P_2 . For each choice of P_1 and P_2 , χ^2 is evaluated and after several iterations the doublet (P_1, P_2) corresponding to the minimum value of χ^2 is chosen. Given some initial value for (P_1, P_2) , the fit procedure starts with an iteration loop in order to calculate and minimize χ^2 . In every iteration the parameters are varied in steps equal to 0.1 within a restricted interval. If the previous minimization converged, a search for an additional local minimum is performed. Finally, the definitive minimum is established and the error matrix of the corresponding parameters is calculated. The global fit to the data distribution is shown in Fig. 9.6 and results in a scaling factor for beauty MC of $P_2 = 1.21 \pm 0.17$ i.e. the factor by which the PYTHIA b cross sections needs to be scaled in order to describe the data. (Note that in all the figures, the scaling factor will be called f_b instead of P_2).

Extraction of the beauty fraction in bins of p_T and η

In order to calculate beauty differential cross sections in the variables: transverse momentum p_T and pseudorapidity η of the jet, the fit procedure was repeated in different bins of p_T and η to estimate the beauty fraction in each of them. Figures 9.7 and 9.8 show the results of the fit as well as the bins used for both variables.

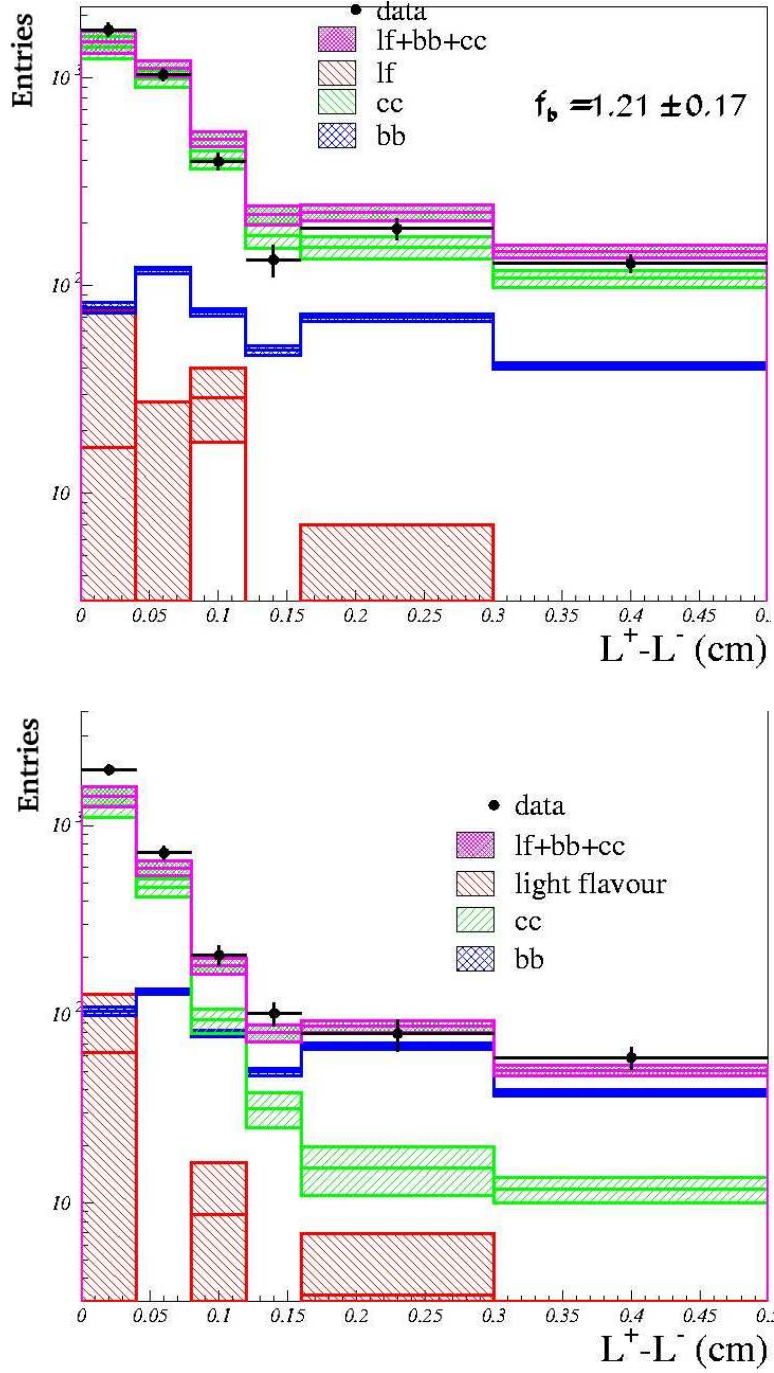


Figure 9.6: Results of global 2-parameters simultaneous fit of $(L_{xy}^+ - L_{xy}^-)$ for the bins: (top) $1.4 \text{ GeV}/c^2 < M_{SV} < 2.0 \text{ GeV}/c^2$ in which the charm contribution dominates and $M_{SV} > 2.0 \text{ GeV}/c^2$ characterized by beauty at high $L_{xy}^+ - L_{xy}^-$ values (bottom). f_b denotes the beauty factor (see text).

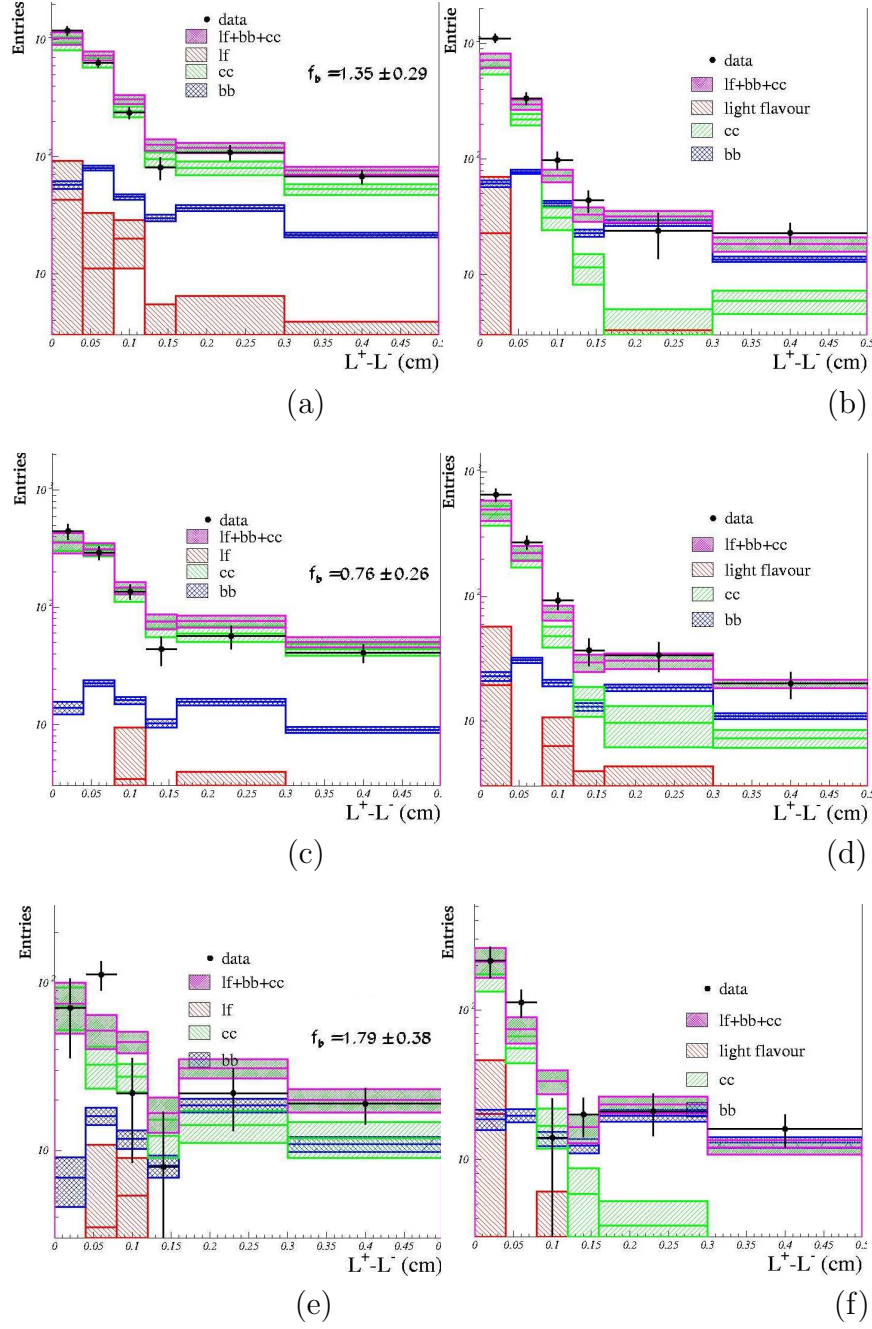


Figure 9.7: $(L^+_{xy} - L^-_{xy})$ distributions in bins of p_T after the fit to the data points. The bins are $7 \leq p_T < 11$ GeV for plots (a)-(b), $11 \leq p_T < 16$ GeV for plots (c)-(d) and $p_T \geq 16$ GeV for plots (e)-(f). The plots on the left side correspond to the distributions in the second bin of the secondary vertex mass, the plots on the right side to the third bin of the secondary vertex mass.

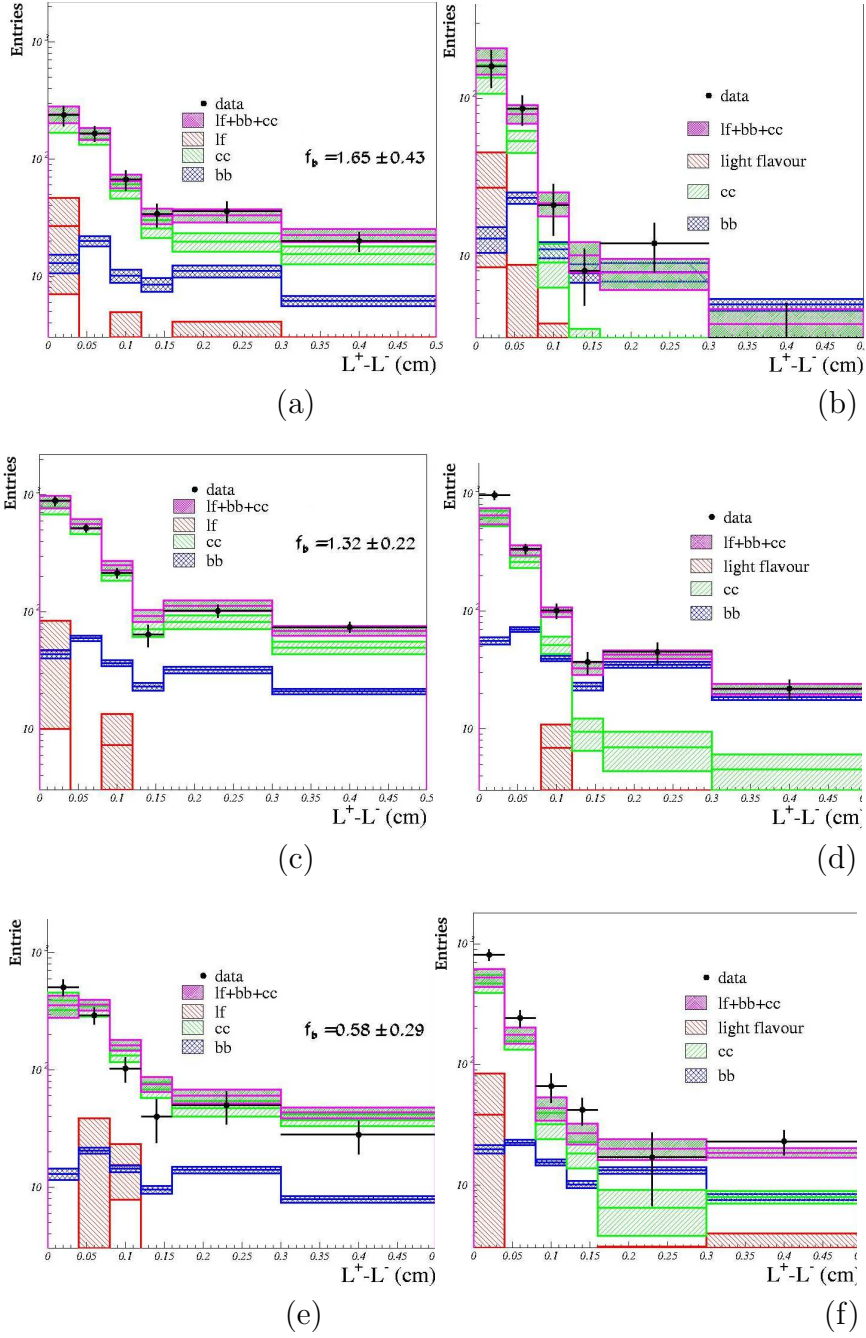


Figure 9.8: $(L_{xy}^+ - L_{xy}^-)$ distributions in bins of η after the fit to the data points. The bins are $-1.6 \leq \eta < -0.6$ for plots (a)-(b), $-0.6 \leq \eta < 0.4$ GeV for plots (c)-(d) and $0.4 \leq \eta < 1.4$ GeV for plots (e)-(f). The plots on the left side correspond to the distributions in the second bin of the secondary vertex mass, the plots on the right side to the third bin of the secondary vertex mass.

Chapter 10

Beauty Quark Cross Sections

This chapter describes the measurement of the inclusive $b\bar{b}$ jet production cross section. Measurements of the differential cross sections with respect to the transverse momentum p_T and pseudorapidity η of the jet are also described. After the systematic uncertainties affecting the measurements are analyzed, the results are compared to previous measurements as well as to NLO QCD predictions and the LO plus parton shower MC predictions of the PYTHIA generator. Finally, a summary and discussion of the results is given.

10.1 Definition of the cross section

The cross section σ of a given process is defined as the number of events in the sample from this process N_{events} per integrated luminosity \mathcal{L} :

$$\sigma_{process} = N_{events}/\mathcal{L} \quad (10.1)$$

In order to measure the visible cross section for beauty inclusive production, the formula can be rewritten as:

$$\sigma_{ep \rightarrow e b \bar{b} X} = \frac{N_b^{rec} \cdot f_b}{\mathcal{L} \cdot \mathcal{A}} \quad (10.2)$$

where N_b^{rec} denotes the number of MC reconstructed b events after the event selection normalized to the data luminosity and f_b is the beauty scale factor obtained from the $(L_{xy}^+ - L_{xy}^-)$ fit (see section 9.5). The quantity $(N_b^{rec} \cdot f_b)$ denotes the number of b events in the data sample.

The acceptance \mathcal{A} , determined from the MC simulation, is defined as the

ratio between the number of b events reconstructed in the detector and those generated N^{true} in the kinematic region of interest:

$$\mathcal{A} = N_b^{rec}/N_b^{true} \quad (10.3)$$

The beauty cross section is calculated within the following kinematic region:

Photoproduction regime (see Sec. 8.1.3):

$Q^2 < 1 \text{ GeV}^2$ and $0.2 < y < 0.8$, selecting two jets with:

$$P_T^{jet1,2} > 7, 6 \text{ GeV} \text{ and } |\eta^{jet1,2}| < 2.5.$$

The measured cross section reported here is an inclusive dijet cross section in $b\bar{b}$ events $\sigma(ep \rightarrow e b\bar{b}X \rightarrow e jjX')$. A b -jet is defined as a jet containing a B hadron, including all b baryons and excitations. The total measured cross section for inclusive beauty production in this kinematic range is:

$$\sigma_{tot}^{measured}(ep \rightarrow e b\bar{b}X \rightarrow e jjX') = 1098. \pm 154. \text{ (stat.) } {}^{+212.}_{-128.} \text{ (syst.) pb.} \quad (10.4)$$

The measurement can be compared to the NLO QCD prediction [59] given below. A good agreement within the still large errors is found:

$$\sigma_{tot}^{NLO} = 836. {}^{+173.}_{-253.} \text{ pb.} \quad (10.5)$$

10.2 Differential cross sections $d\sigma/dp_T^{b-jet}$ and $d\sigma/d\eta^{b-jet}$

The differential cross sections calculated as a function of transverse momentum $d\sigma/dp_T^{b-jet}$ and pseudorapidity $d\sigma/d\eta^{b-jet}$ of the b -jet are shown in Table 10.1 and 10.2 as well as in Figure 10.1 (a) and (b) respectively.

The measured differential cross sections are compared with a previous ZEUS measurement [33] which uses a completely different event selection and beauty extraction method (see Sec. 3.3.1). Specifically, it performed a study on beauty photoproduction with a integrated luminosity of 110 pb^{-1} selecting events which contain at least two jets and a muon. The beauty signal was extracted by exploiting the distribution of the transverse momentum of the muon relative to the axis of the jet in which it is included (known as p_T^{rel} method). The measured μ -jet cross sections were then corrected to obtain the cross sections for b -jets in dijet events by using PYTHIA and accounts for the $b \rightarrow \mu$ branching ratio. The two independent measurements are consistent and in good agreement within errors.

A direct comparison of the results and the NLO QCD predictions is also performed. The NLO calculations are based on the FMNR program and were corrected for hadronization effects as explained in section 2.8.2. As described there, the main systematic uncertainties on the NLO calculations come from the b -quark mass and the renormalization and factorization scales. Though the statistical and systematic errors are high for the 2004 data, the differential cross sections are in a reasonable agreement with the NLO QCD predictions.

Finally the measured cross sections are compared to the absolute predictions of the MC model PYTHIA. The MC reproduces reasonably the shape of the data and the normalization within the statistical and systematic errors. The differential cross sections are therefore compatible with the PYTHIA model based on leading order calculations plus parton showering.

p_T^{b-jet}	$d\sigma/dp_T^{b-jet} \pm stat. \pm syst.$	published ZEUS result
$7 \leq p_T < 11 \text{ GeV}$	$187 \pm 40^{+55}_{-13}$	$137 \pm 14^{+21}_{-27}$
$11 \leq p_T < 16 \text{ GeV}$	$25.3 \pm 8.6^{+5.9}_{-5.2}$	$43.8 \pm 5.5^{+7.7}_{-12.0}$
$p_T \geq 16 \text{ GeV}$	$6.0 \pm 1.3^{+0.5}_{-0.2}$	$5.7 \pm 1.2^{+1.03}_{-0.9}$

Table 10.1: Differential cross section for b-jets as a function of p_T^{b-jet} . For comparison, the previous ZEUS published results [33] are also quoted.

η^{b-jet}	$d\sigma/d\eta^{b-jet} \pm stat. \pm syst.$	published ZEUS result
$-1.6 \leq \eta < -0.6$	$235 \pm 61^{+1}_{-22}$	$152 \pm 29^{+24}_{-31}$
$-0.6 \leq \eta < 0.4$	$438 \pm 73^{+46}_{-43}$	$356 \pm 41^{+59}_{-53}$
$0.4 \leq \eta < 1.4$	$189 \pm 94^{+146}_{-30}$	$275 \pm 45^{+53}_{-73}$

Table 10.2: Differential cross section for b-jets as a function of η^{b-jet} . For comparison, the previous ZEUS published results [33] are also quoted. The differential cross section $d\sigma/d\eta^{b-jet}$ is in the previous measurement calculated as a function of four bins. The last one, corresponding to $1.4 \leq \eta < 2.5$, was rejected in this analysis because of lack of statistics that made the fit impossible to converge.

10.3 Systematic errors

The main systematic uncertainties affecting the cross section measurement are due to the event reconstruction and extraction method used to determine the beauty signal. A list of the different systematic checks and their effects on the total measurement is given below and summarized in table 10.3. For

the differential cross sections the systematic errors are calculated for each bin and are included in all the results and figures.

- **M_{SV} reweighting:** in Chapter 9 the correction applied to the secondary vertex mass M_{SV} shape of MC was described. The cross sections were recalculated not applying the correction at all. The effect on the total beauty cross section is +2.5%.

- **Fit systematics:** to check the reliability of method of extraction of the beauty signal, all the fits have been redone varying different parameters:

- **$L_{xy}^+ - L_{xy}^-$ shape correction for charm:** this correction was explained in 9.4 and it is evaluated recalculating the cross sections not applying the reweighting factor $[data/MC]_{charm}$ at all. This causes a variation of the total cross section of -11.5%.
- **First bin:** despite the correction applied in order to get a good MC description of the $(L_{xy}^+ - L_{xy}^-)$ distributions, the lowest bin of the $(L_{xy}^+ - L_{xy}^-)$ plot in the mass region $M_{SV} > 2. GeV/c^2$ (see f.e. Fig. 9.6 (b)) shows still a noticable disagreement between data and MC. The systematic error caused by this discrepancy was calculated fitting the distributions again without this first bin. The biggest systematic uncertainty arises from this, leading to a variation of the cross section of +15.8%.
- **One fit:** The last systematic check concerning the fit method was done by fitting only one distribution instead of two. To perform this check, the $(L_{xy}^+ - L_{xy}^-)$ distribution for $M_{SV} > 2. GeV/c^2$ was chosen since it is the one in which the beauty and charm distribution shapes are different enough to carry out a fit successfully. The variation of the cross section was found to be -1.6%.

- **Luminosity correction:** The integrated luminosity collected by the ZEUS experiment during the 2004 data taking period is known with an error of $\pm 4.2\%$ [52], which leads to systematic uncertainties of $\pm 4.2\%$.

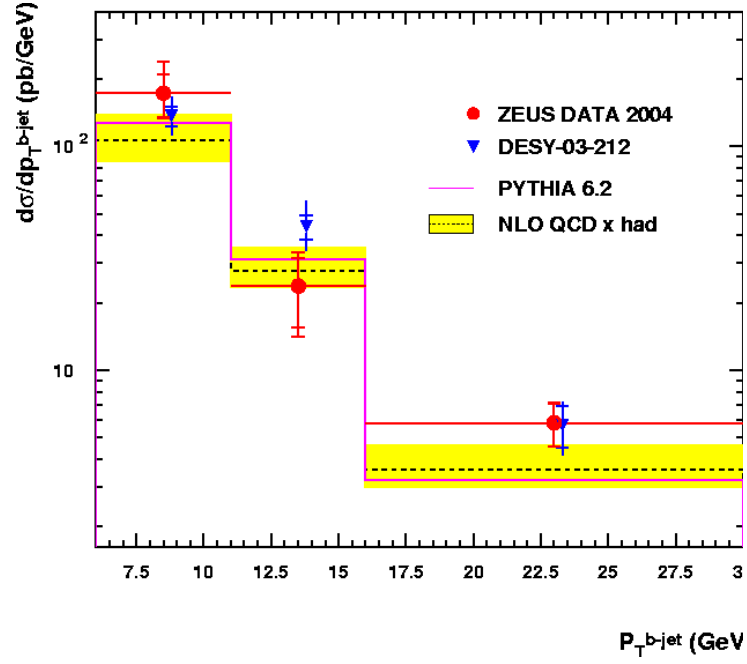
- **FLT track acceptance correction:** during 2004 the First Level Trigger track veto efficiency was not properly described in the Monte Carlo since the CTD was operated at 95% of its HV nominal value. A rough estimate of 1.05 ± 0.05 was obtained from independent studies [77] corresponding to a cross section variation of $^{+10\%}_{-0\%}$.

The total systematic error is calculated as the sum of all the individual uncertainties in quadrature in the total and in the differential cross sections.

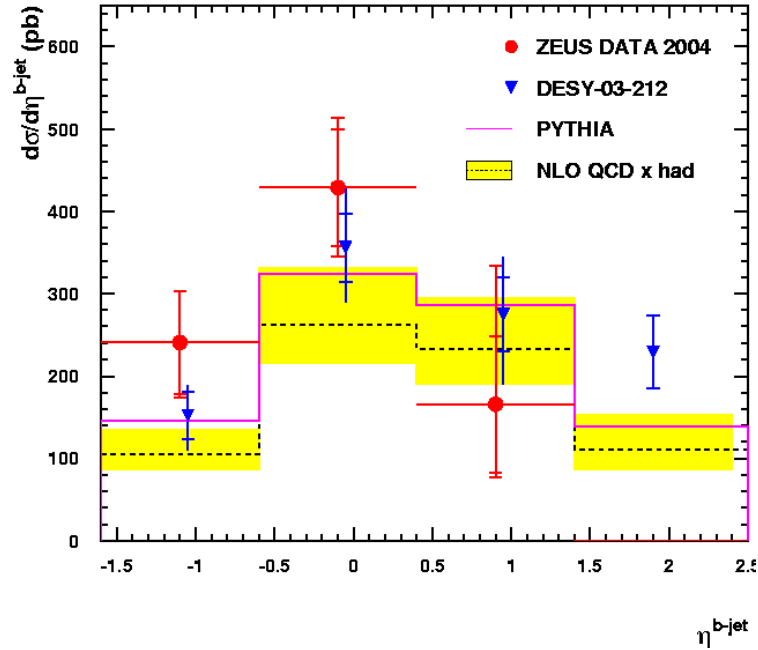
The calculation is done separately for the positive and negative variations. This analysis is dominated by the statistical errors. Thus, other systematic checks such as the uncertainties due to jet energy scale or PHP selection are considered small compared to it.

Description of systematic checks	$\Delta\sigma(\%)$
M_{SV} reweighting	+2.5%
Fit systematics:	
$c\bar{c}$ -distribution reweighting	-11.6%
Fit w/o first bin	+15.8%
Only one 2-param fit (3^{rd} mass bin)	-1.6%
Luminosity measurement $\pm 4.2\%$	$\pm 4.2\%$
FLT track accep. 1.05 ± 0.05	$+10\%$ -0%
Total	+19.3% -11.7%

Table 10.3: Summary of the systematic errors for the total cross section measurement



(a)



(b)

Figure 10.1: Differential cross sections as a function of (a) the transverse momentum p_T^{b-jet} and (b) the pseudorapidity η^{b-jet} of the jet. The data (red circles) are compared to the predictions of PYTHIA (solid line) and to a previous ZEUS measurement [33] (blue triangles moved slightly to the right for a better comparison). The full error bars are the quadratic sum of statistical (inner) and systematic uncertainties. The data are also compared to the predictions of NLO QCD (dashed line histogram) corrected to hadron level. The band around the NLO prediction represents the uncertainty on the theoretical prediction (see Sec. 2.8.2)

Chapter 11

Summary and Conclusions of 2004 Data Analysis

For the first time in ZEUS, inclusive beauty photoproduction has been measured directly using inclusive dijet events: $\sigma(ep \rightarrow ebb\bar{b}X \rightarrow ejjX')$. The feasibility of a new b -tagging method, the secondary vertex algorithm, never used before in this experiment, was explored and its applicability verified. Total visible and differential cross sections of this process were measured in the kinematic region defined by $Q^2 < 1 \text{ GeV}^2$ and $0.2 < y < 0.8$ in events with two or more jets of transverse momentum $p_T^{jet1,2} > 7(6) \text{ GeV}$ in the central range of pseudorapidity $|\eta_{jet1,2}| < 2.5$. This measurement used a data sample of approximately $\mathcal{L} \sim 35 \text{ pb}^{-1}$ collected by the ZEUS detector in the year 2004 after the upgrade of HERA. During this time HERA operated with a center of mass energy of $\sqrt{s} = 320 \text{ GeV}$.

The fraction of events from beauty quark production was extracted on a statistical basis using the decay length and the invariant mass at the secondary vertex. The measured differential cross sections were compared to the MC model PYTHIA, based on leading order plus parton shower calculations, and to an NLO QCD prediction based on the FMNR program corrected for hadronization effects. Both predictions were able to give a good description of the differential distributions within the still large statistical and systematic errors.

The measurements were also compared to a previously published beauty photoproduction result obtained using decays into muons in dijet events with a HERA I dataset of an integrated luminosity of 110 pb^{-1} . μ -jet cross sections were measured and then corrected to obtain the b -jet cross sections. Both results were found to be in good agreement within errors.

The statistical and the systematic errors on the measurements are still high. The higher luminosity of the complete HERA II dataset (for 2005 data

only, the luminosity is already approximately 4 times higher than for 2004) will reduce the statistical uncertainties. Also the systematic errors are considerably improved with the use of the new available data (see next Chapter). A better understanding of the traditional tracking and vertexing techniques as well as the implementation of new approaches are expected to turn this exploratory analysis into a much more precise measurement.

As a final conclusion from this analysis, figure 11.1 shows the b -quark differential cross section as a function of the quark transverse momentum, $d\sigma(ep \rightarrow bX)/dp_T^b$. It was obtained extrapolating the cross section for dijet events using the NLO QCD prediction corrected for hadronization. The \bar{b} quark was not considered in the definition of the b -quark cross section. This extrapolation allows the direct comparison not only with previous ZEUS measurements given at the b -quark level but also with H1 results [32] (see chapter 3.3). The independent measurements are consistent and in agreement with the NLO QCD predictions. However there is a light tendency of the data to lie above the theoretical predictions, especially for lower p_T^b .

Summarizing, the goal to establish a new beauty tagging algorithm (independent of the ones used so far in ZEUS) and to exploit its full potential obtaining a first measurement of inclusive beauty photoproduction in dijets events was reached. A precision measurement clearly went beyond the scope of this exploratory analysis, however it confirms the feasibility of the method and provides the basic knowledge on which to build soon a precision analysis with the new data.

HERA

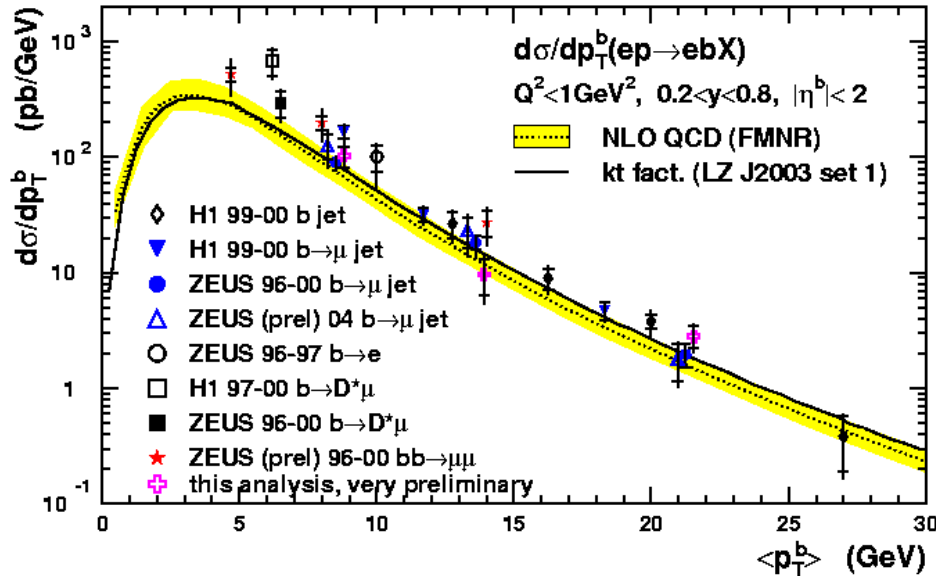


Figure 11.1: Differential cross sections for b -quark as a function of the b -quark transverse momentum p_T^b for b -quark pseudorapidity $|\eta^b| < 2$ and for $Q^2 < 1 \text{ GeV}$ and $0.2 < y < 0.8$. The pink crosses are the results obtained in this analysis with the ZEUS 2004 data. The filled circles show the ZEUS published results which use the p_T^{rel} tagging method in dijet plus muon events. Also other results from ZEUS and H1 are shown [32]. The dashed line is the NLO QCD prediction with its theoretical uncertainty (shaded band).

Chapter 12

2005 Data Analysis - Preliminary Results

This chapter gives an outlook into the analysis of the data taken during the end of 2004 and the year 2005 (called 2005 data for short) which is still ongoing. The studies performed so far as well as the first preliminary results are shown here. As already mentioned previously, this new analysis is strongly motivated by the results obtained in the exploratory analysis of the 2004 data presented in this thesis and by the remarkable improvements of the ZEUS tracking software and the detector alignment and calibration since the first analysis. An immediate consequence of those improvements is e.g. the proper treatment of the errors which is translated into the capability to determine the decay length significance L_{xy}/σ_{xy} . A more accurate alignment of the MVD with respect to the other detector components also contributes to a higher precision of the measurements. The chapter is ordered following the same structure as presented for the 2004 data analysis. Only the features new to this analysis will be explained in detail whereas the procedures and reconstruction methods common with the 2004 analysis are only shortly reminded. The results presented here are very preliminary.

Concerning the detector status, the STT component (see Sec. 4.2) was turned off during the 2005 data taking period because of technical problems. However, this does not affect this analysis since the STT was not yet used in the early reconstruction of the 2004 data.

12.1 Track reconstruction

The package used for the reconstruction of the tracks in this analysis is an improved version of the KFFIT package used in the 2004 analysis, already

described in detail in section 6.1. The input track parameters and covariance matrices used initially by KFFIT were not treated correctly, and the algorithm had to be modified accordingly. Moreover, some parts of the code were rewritten to achieve further improvements such as better momentum resolution, faster computing time and proper calculation of the errors on the track parameters.

After all improvements [90], an improvement on the momentum resolution by a factor around 1.4 has been reached without affecting resolutions on the other track parameters. The estimated errors on the track parameters have been corrected by a factor 1.2 and the computing time has been reduced by a factor 8.

12.2 Secondary vertex reconstruction

While for the 2004 data analysis the routine VXU (see Sec. 7.4) was used to reconstruct the secondary vertices, the package `vxLite` is chosen for reconstruction of secondary vertices in the 2005 data analysis. `vxLite` is the vertexing package included in the newly developed `tLite` library [50], a collection of tracking related analysis tools. The routine `vxLite`, like VXU, is an implementation of the global vertex fit algorithm by Billoir and Fröhlich already explained in the Chapter 7. The main and nearly only difference of `vxLite` with respect to VXU¹ relies on the intrinsic cut on the χ^2 of the vertex. While VXU rejected automatically all tracks which lead to a vertex with $\chi^2 > 100.$, the `vxLite` output is a fitted vertex with all the tracks specified by the user. A dedicated study about the effects of both routines on the particular case of beauty tagging using secondary vertices was performed in [63]. The study is based on the comparisons of the reconstructed transverse decay length. As expected, both routines reproduce the same results with the exception of the high χ^2 vertices not removed by `vxLite`. Nevertheless, the rejection of such bad vertices can be performed at analysis level by cutting explicitly on χ^2 . The tracks to be fitted were selected and associated to jets using the same requirements than in the 2004 analysis (see section 8.3) except for one: in the first analysis, all tracks which led to a vertex made of two oppositely charged tracks with mass compatible with K_s^0 and Λ were removed. These tracks are not rejected in the 2005 analysis since the explicit cut on χ^2 removes the bias from K_s^0/Λ candidates, characterized by long positive decay length values. This will be described in section 12.6.

¹Another difference is the programming language. The routine `vxLite` uses C++ while VXU uses FORTRAN

As in the first analysis, the primary vertex values, calculated using the fast vertex fitting algorithm (see Sec. 7.3), were replaced in each event by the beam spot. For the 2005 data sample, instead of calculating the beam spot as an average from the primary vertex position over all the events in the same run, the average was done for each sequence of 2000 data events in order to achieve better accuracy for its time dependence.

The choice of using the `vxLite` routine over `VXU` is justified by convenience purposes. First, `tlite` is the package supported officially by the ZEUS tracking group. Second, in contrast to `VXU` the `vxLite` routine is more suitable to be integrated within the standard ZEUS environment. Moreover, `vxLite` has been successfully implemented within the ZEUS common analysis software `ORANGE` as well as within the ZEUS event display `ZeVis`. A short overview about the latter is given below.

12.3 Event Visualization with ZeVis

Event displays are very important tools for the analysis of high energy particle collisions. The visualisation in a graphical way of the event and the detector geometry contributes to a major understanding of the physics of the recorded interaction as well as to the verification and optimization of the algorithms applied in the reconstruction of event objects. Therefore, the `vxLite` version of the secondary vertices algorithm was implemented in the ZEUS event display (`ZeVis`) making possible the visualisation of the analysis tracks (tracks fitted to the secondary vertices) and the reconstructed vertices.

Figure 12.1 shows the event display of a beauty candidate ($\gamma g \rightarrow b\bar{b}$ with $b \rightarrow \mu X$, $b \rightarrow J/\Psi X$ and $J/\Psi \rightarrow \mu\mu$) in different views. The region around the primary vertex before and after calling the secondary vertex algorithm is shown. The algorithm not only finds a secondary vertex but also allows the estimation of the new primary vertex ("reduced" primary vertex) in which the tracks which has been already fitted to a secondary vertex are removed from the fit.

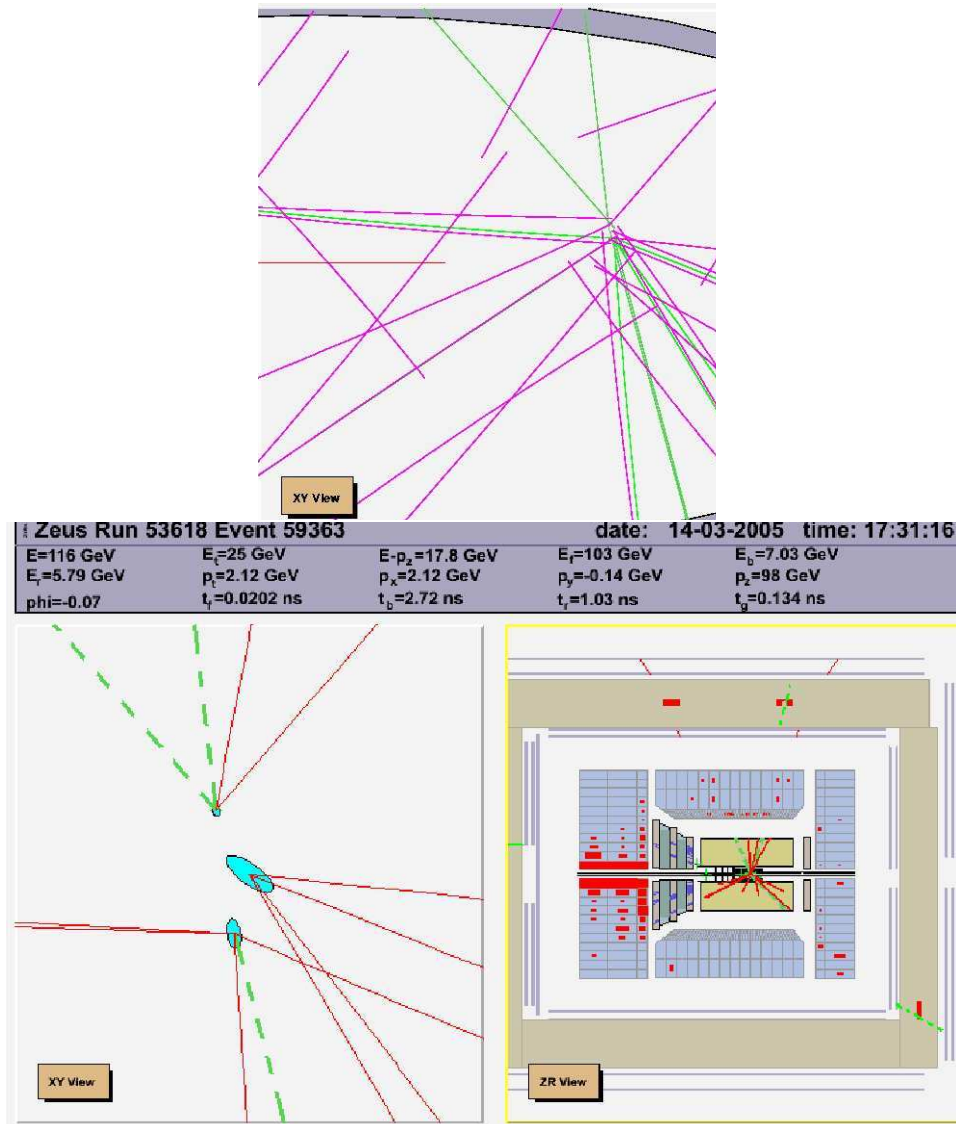


Figure 12.1: Display of a trimuon beauty candidate. The top figure shows the interaction region before the secondary vertex algorithm is called. At this stage no selection of tracks has been performed yet. In the bottom figure are depicted the primary interaction region (left) where the secondary vertices found with the secondary vertex algorithm can be seen and the complete event in the ZR view (right). The green (dashed lines) tracks correspond to muon candidates. The ellipses on the top figure represent the ellipse shape of the beam pipe.

12.4 Event Selection

The dataset used for this analysis was collected by the ZEUS detector in the running period covering the end of 2004 and 2005. It corresponds to a luminosity of 132.6 pb^{-1} (approximately 3.5 times higher than the luminosity of the 2004 dataset). The same data quality flags, EVTAK and MVDTAK (see Sec. 8.1.1) as in the 2004 data analysis, have been applied here.

The MC samples were obtained using the PYTHIA MC generator in the photoproduction regime ($Q^2 < 1$) and are given in Tab. 12.1. Both, the data- and MC-ntuples, were produced with the software package ORANGE within the frame of the “common ZEUS ntuple” project². This is the first analysis using the common ntuples. The Ntuples were completed few weeks before this thesis was submitted

The trigger and PHP selection applied are exactly the same as for the 2004 data analysis.

Process	$\mathcal{L}[\text{pb}^{-1}]$	Process	$\mathcal{L}[\text{pb}^{-1}]$
$b\bar{b}$ direct	1303.91	inclusive lf. direct	$\sim 112.$
$b\bar{b}$ resolved	1290.01	inclusive lf. resolved	$\sim 112.$
$b\bar{b}$ excitation in γ	1290.00	quark excitations contained in the resolved sample	
$b\bar{b}$ excitation in p	1289.96		
$Q^2 < 1 \text{ GeV}^2$		$Q^2 < 1 \text{ GeV}^2$ and $P_{Tmin} > 2 \text{ GeV}/c$	
not cuts		2 jets both jets $E_T > 4$ and $-3. < \eta < 3.$	

Table 12.1: The MC samples used for the 2005 data analysis. All samples were obtained using the PYTHIA MC generator. Inclusive lf. samples refers to processes producing both light and heavy flavours.

12.5 Control distributions

The control distributions, shown in this section (Fig 12.2 to 12.4), illustrate the good description by the MC simulation of most of the track and jet related variables. All the MC distributions were normalized to the corresponding data areas. Displayed are all relevant variables after the trigger

²A “common ZEUS ntuple” has been designed and partially produced to provide a convenient and worthwhile analysis of the HERA data beyond 2010 when the software infrastructure is not longer guaranteed. Likewise, the common ZEUS ntuple is intended to facilitate the comparisons between physics analysis and to save considerable amount of time in the production of own ntuples by the physicist. These two advantages makes more efficient the physics analysis, specially under the conditions of manpower lack.

plus photoproduction selection: azimuthal angle, pseudorapidity, transverse momentum and number of jets in the event for all the jets with $p_T > 7 \text{ GeV}/c$ and a fitted secondary vertex. Concerning the track variables, the distributions shown are the azimuthal angle, polar angle, transverse momentum, impact parameter, number of outer superlayers in the CTD and number of hits in the MVD. All distributions show a good agreement except the azimuthal angle for both, jets and tracks, where the MC simulation does not yet describe properly some dead regions found in the MVD.

12.6 Decay length significance

The procedure followed to extract the content of beauty signal in the sample is similar to the one used for the 2004 data analysis, i.e. based on the characteristic spectrum of the transverse decay length L_{xy} . For this analysis L_{xy} has been defined, different than for the 2004 analysis (see section 8.2.4), as the distance between the primary and secondary vertex in the XY plane projected onto the jet axis:

$$L_{xy} = (\vec{S} - \vec{P}) \cdot (\hat{j}/|\hat{j}|) \quad (12.1)$$

where \vec{P} and \vec{S} are the position vectors of the primary and secondary vertices in the XY plane and \hat{j} the vector along the jet axis. The reason for this redefinition, suggested by the tracking experts [70], is purely statistical, the distribution of the so-defined L_{xy} obeys the gaussian statistic. As in the 2004 analysis (see Sec. 8.1.1), the primary vertex refers to the average primary vertex i.e. the beam spot.

For the calculation of the decay length, following secondary vertices were removed in order to reject secondary interactions of no interest:

- all vertices when $|Z_{\text{primary-vertex}}| > 50 \text{ cm}$,
- all secondary vertices at more than 1 cm in the $x - y$ plane and 30 cm in the z plane from the nominal interaction point,
- and all secondary vertices with $\chi^2 > 6$.

While the first two cuts are already known from the 2004 analysis, the last one is new for this analysis and was done in order to discard high χ^2 vertices not removed by `vxLite` automatically (see sec. 12.2). The cut was optimized based on the dedicated study presented in [63] and already mentioned above. Figure 12.5 shows the mirror distribution ($L_{xy}^+ - L_{xy}^-$) before and after the χ^2 cut has been applied. The flat tail of the distributions for the MC charm and

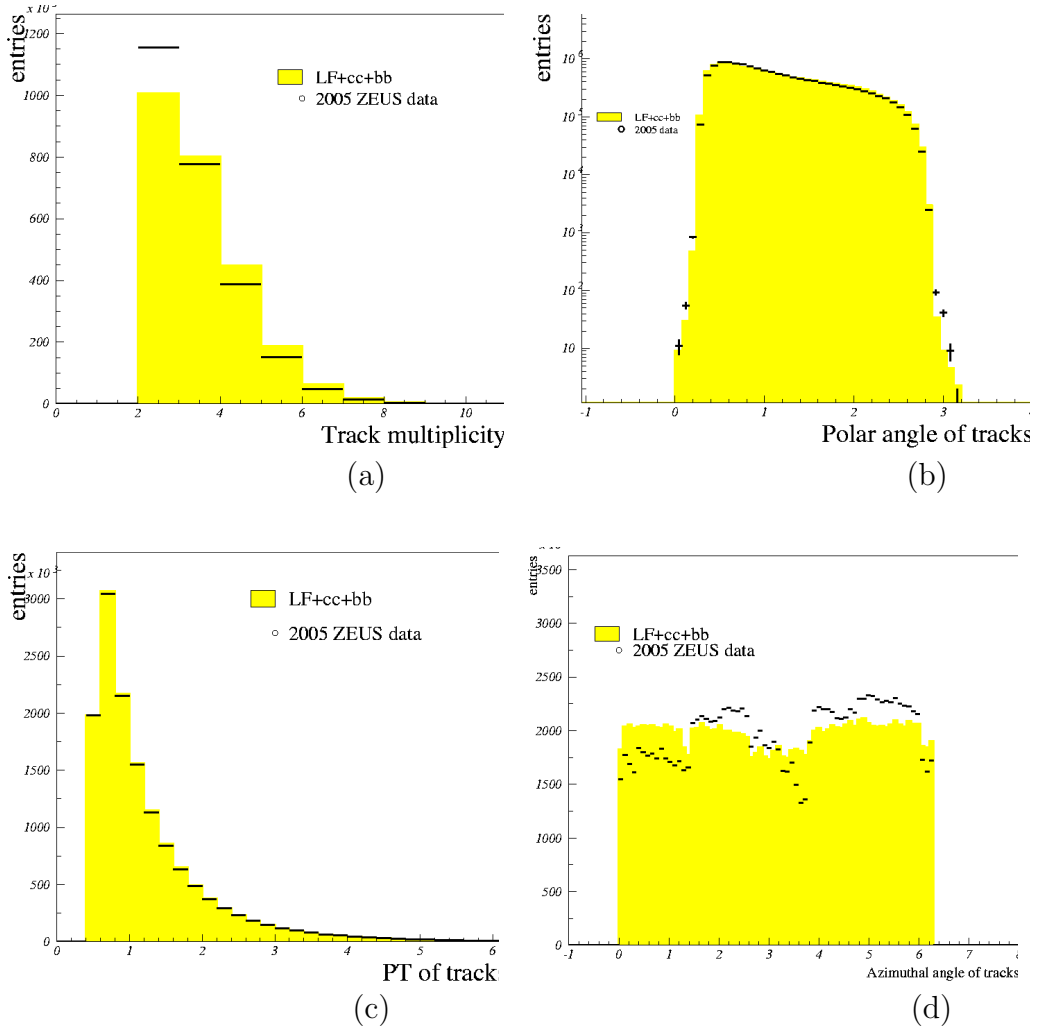


Figure 12.2: Control distributions of the track related variables. Only tracks passing the final selection (see text) are considered here: (a) Number of tracks in the jet, (b) polar angle θ , (c) transverse momentum p_T of the tracks and (d) azimuthal angle ϕ . All the Monte Carlo histograms are normalized to the data area.

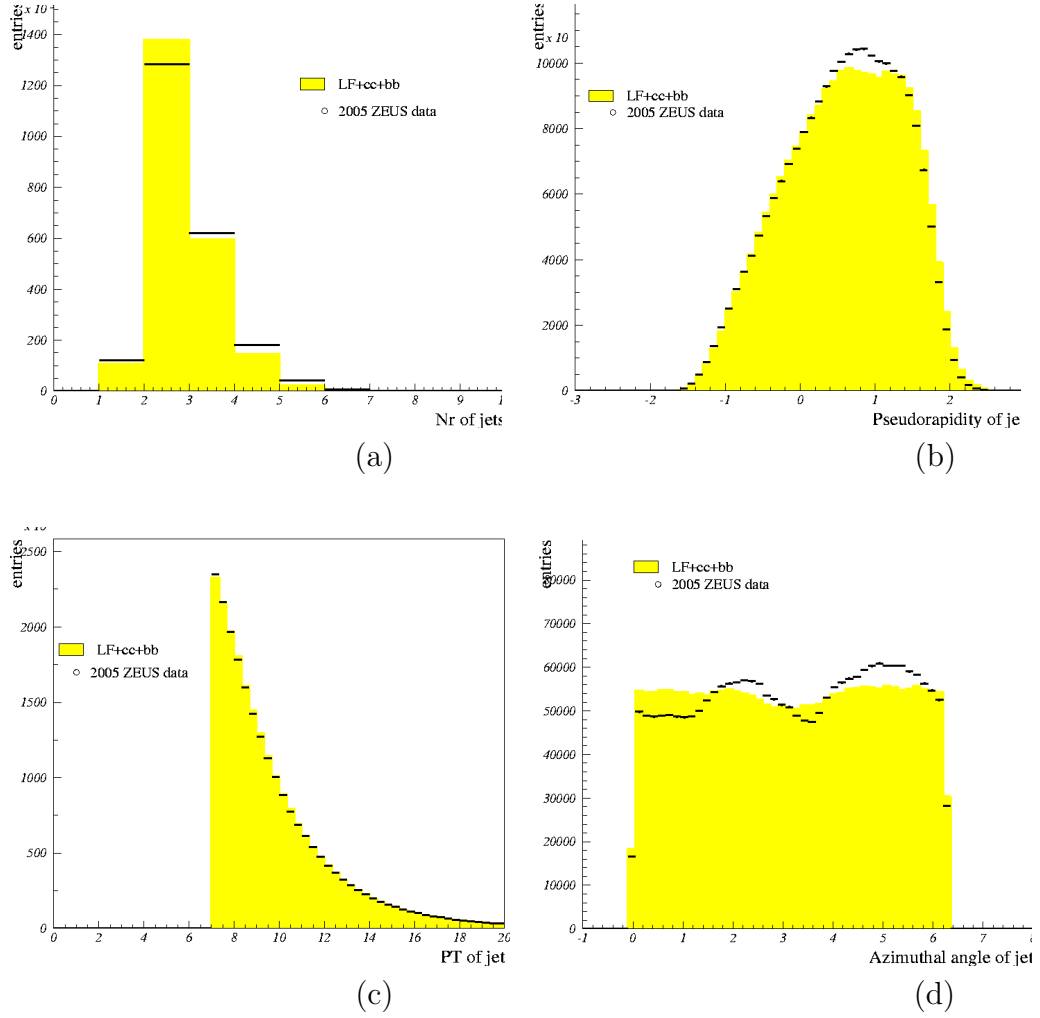
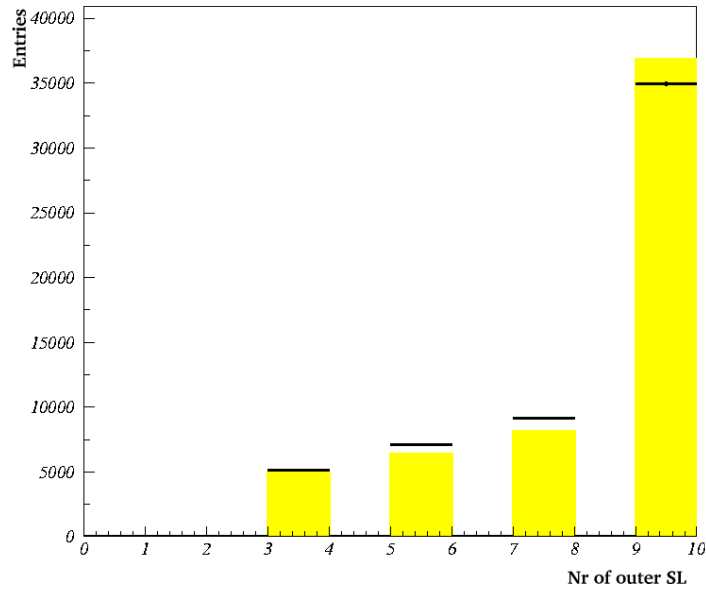
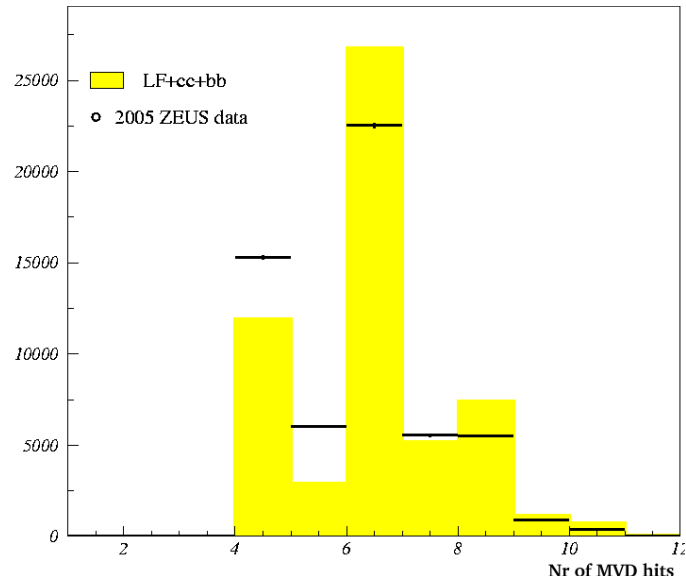


Figure 12.3: Control distributions of the jet related variables: (a) Number of jets in the event, (b) pseudorapidity η_{jet} , (c) transverse momentum p_T of the jets and azimuthal angle ϕ . All the Monte Carlo histograms are normalized to the data area. All the jets with a reconstructed secondary vertex are shown.



(a)



(b)

Figure 12.4: Control distributions of the track variables related to the MVD-CTD:(a) Number of Super Layers in the CTD crossed by the selected tracks, (b) Number of MVD hits used in the reconstruction of the track. All the Monte Carlo histograms are normalized to the data area.

light flavour contributions, corresponding mostly to long-lived particles such as K_s^0 and Λ , drops rapidly to zero after the χ^2 cut has been required. This is due to the wrong association of tracks corresponding to K_s^0/Λ candidates to the jets to be fitted, and leads to a poor reconstruction (high χ^2) of the secondary vertex.

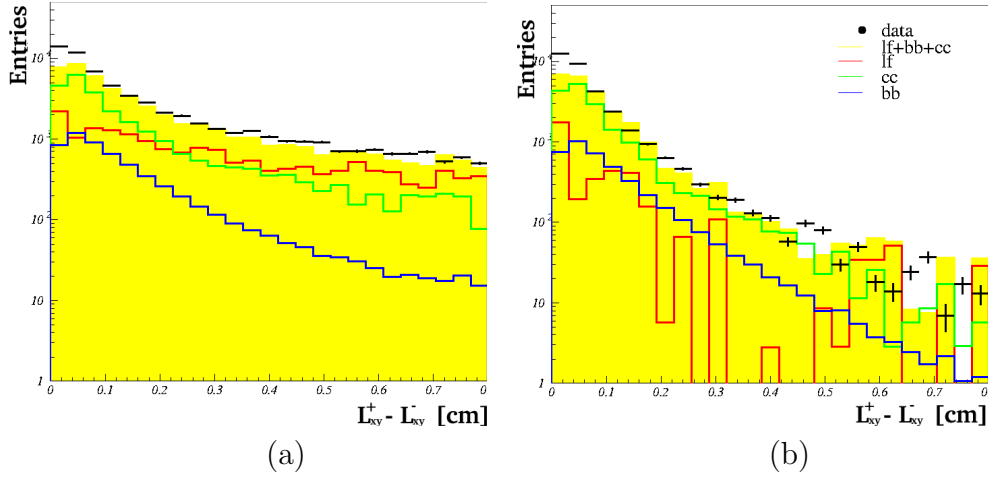


Figure 12.5: $(L_{xy}^+ - L_{xy}^-)$ - distribution before (a) and after the $\chi^2 < 6$ cut has been applied (b).

The biggest improvement of this analysis with respect to the 2004 analysis relies however on the use of the transverse decay length significance rather than the decay length itself to obtain the fraction of beauty. The significance S is defined as $S = L_{xy}/\sigma_{L_{xy}}$ where $\sigma_{L_{xy}}$ denotes the error on L_{xy} and is computed from the covariant matrix elements of the primary and secondary vertex positions. It should be remembered that the significance could not be used for the 2004 analysis due to the wrong treatment given to the errors in the tracking procedure (see section 8.2.4).

The reliability of the method depends on a good MC description of all variables used directly or indirectly in the calculation of the significance S (see control plots in last section 12.5) as well as the S -distribution itself. The later implies a good description of the decay length L_{xy} distribution and the its error $\sigma_{L_{xy}}$, both separately. However, and despite the improved tracking/vertexing techniques with respect to 2004 data, a comparison of these distributions between data and MC simulations shows important disagreements. In figure 12.6 the significance distribution as well as the decay length and the $\sigma_{L_{xy}}$ spectrum are depicted separately. It can be seen that, while the significance distribution shows a reasonable agreement between data and MC, the decay length and the errors are badly described by the MC simula-

tions.

This disagreement has two components:

- the data distribution is significantly wider than the MC distribution,
- and the data is slightly shifted to the right with respect to the MC distribution.

Both components were handled individually as described in the following subsections. While the first disagreement was corrected by applying a MC smearing, the second disagreement is still under investigation.

12.6.1 MC smearing

As it has been seen, the decay length distribution is significantly wider than the MC distribution while this problem is not present in the significance distribution. In order to achieve a satisfactory agreement between data and MC in the decay length distributions (and error) without impairing the agreement in the significance distribution, a smearing procedure³ suggested by the ZEUS tracking experts [71] and motivated by possible misalignments and the limited resolution of the micro vertex detector, was performed.

The procedure consists in smearing the mean value of the MVD hit clusters and its error. The study is based on a subsample of the 2005 data of approximately $L \sim 17.24 \text{ pb}^{-1}$ and inclusive light flavour MC sample $L \sim 8.4 \text{ pb}^{-1}$. The collection of routines used for this procedure as well as the cross checks carried out by other colleagues can be found in [91]. The main idea of this smearing consists of adding to the mean value of the MVD hit a randomly generated number multiplied by an empiric value. This empiric value is expressed in units of the pitch of the MVD ($120 \text{ } \mu\text{m}$):

$$\langle x \rangle_{\text{smeared}} = \langle x \rangle + N_{\text{RAN}} \cdot \frac{y}{120}. \quad (12.2)$$

where $\langle x \rangle$ is the mean value of the MVD cluster, N_{RAN} denotes a random number generated in a Gaussian distribution of mean zero and variance one and y is the empiric smearing value. The error of the MVD cluster mean was accordingly smeared by adding the empiric smearing value in quadrature.

The smearing value y was optimized by comparing MC distributions obtained for different y values (from $20 \text{ } \mu\text{m}$ to $60 \text{ } \mu\text{m}$, in steps of $10 \text{ } \mu\text{m}$) with the data. The smearing value of $y = 40 \text{ } \mu\text{m}$ was found to be the best fit to

³The smearing procedure used in the 2004 analysis was based on single tracks and thus, is not optimal for an analysis based on the decay length significance

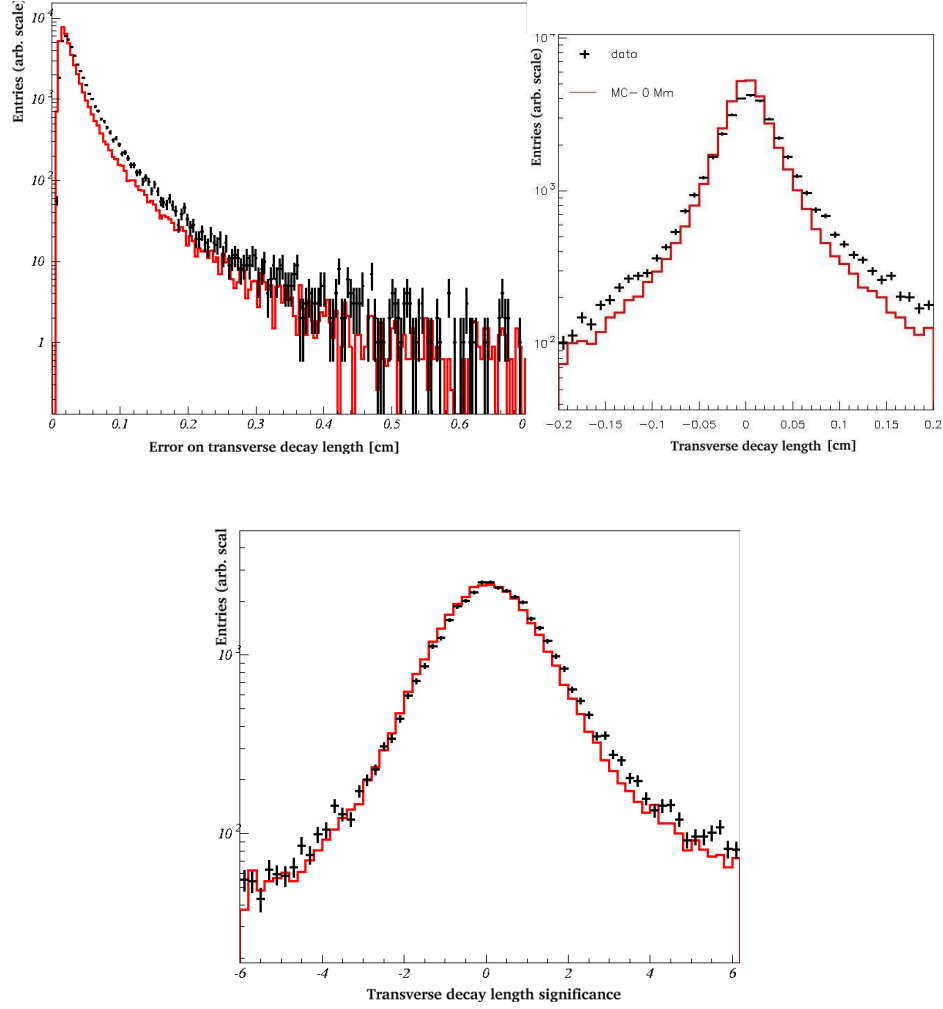


Figure 12.6: The decay length significance $S = L_{xy}/\sigma_{xy}$ (down), transverse decay length L_{xy} (top right) and error distribution σ_{xy} (top left) for MC simulations (solid line) compared to the data (black crosses). Shown are only secondary vertices associated to the highest- P_T jet in the event. The data distribution is significantly than the Monte Carlo distribution.

the data⁴ (see Fig. 12.7). Different checks were performed to confirm that the smearing of the MVD hits does not depend on the azimuthal angle, the transverse momentum or the total charge of the secondary vertices. Fig 12.9 and 12.10 show the distributions of the decay length and the decay length significance in different bins of the azimuthal angle ϕ . The bins have been chosen in the way that the MVD areas covered by 2-cylinders and 3-cylinders of ladders can be observed separately (see Fig. 12.8). The goal followed was to discard/establish a dependence on the azimuthal angle coverage of the smearing procedure as well as of the calculation of the decay length itself. No significant dependence was observed. Fig 12.11 show similar distributions in bins of the average transverse momentum in the secondary vertex. A significant dependence of the smearing procedure on p_T is not found either.

One should notice that smearing the MVD hits by $40\ \mu m$ is not justifiable any more within the limits of the accepted misalignment and detector resolution effects (expected resolution $20 - 30\ \mu m$ [71]). However, as seen in the figures, the improvement of the decay length and decay length significance MC distributions is remarkable which leads to the conclusion that this kind of smearing is obviously taking account other effects which are not yet understood but are clearly present in the data.

⁴In the study shown here, the best estimate was found to be $50\ \mu m$. Differences between an smearing of $40\ \mu m$ and $50\ \mu m$ were found to be insignificant. Since other independent studies [83, 82] found that the value of $40\ \mu m$ fit better the data not only for decay length distributions but also impact parameter distributions, the physics group agreed to take this value as the best estimate

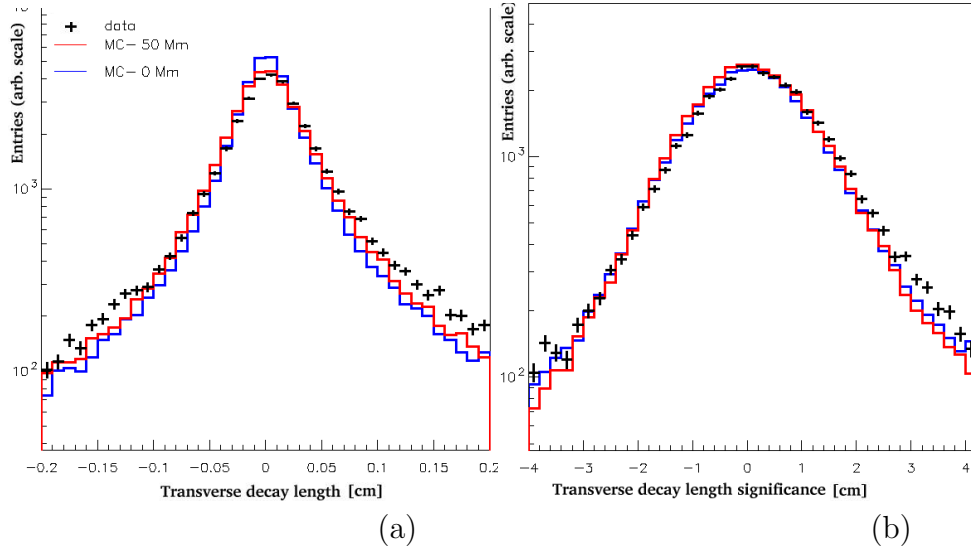


Figure 12.7: The MC transverse decay length (a) and decay length significance distribution (b) compared to the data before and after a MVD hits smearing of $50 \mu\text{m}$ has been performed. Shown are only secondary vertices associated to the highest- P_T jet in the event. The agreement between MC and data distributions is significantly improved after the smearing.

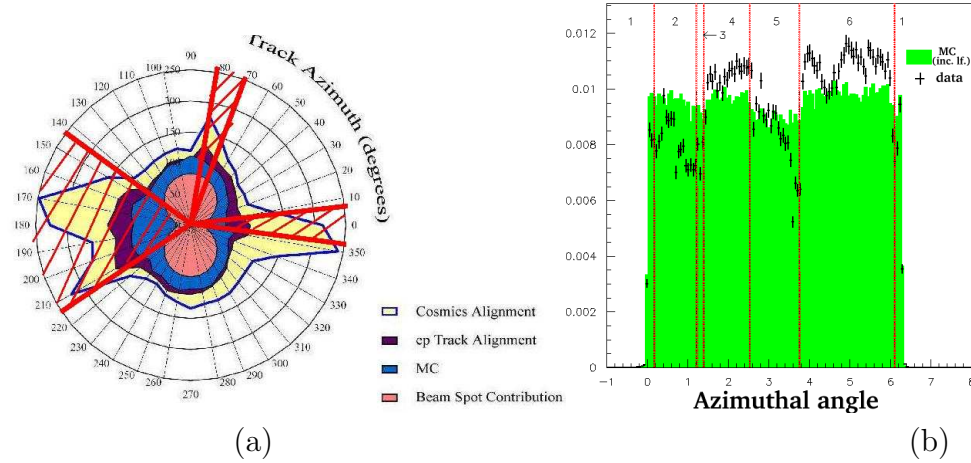


Figure 12.8: (a) Radar plot of MVD alignment. The hatched contours indicated the areas covered only by 2 cylinder of ladders. (b) Distribution of the azimuthal angle of all tracks fitted to a secondary vertex. The distribution is divide in zones corresponding to the areas shown in Fig.(a)

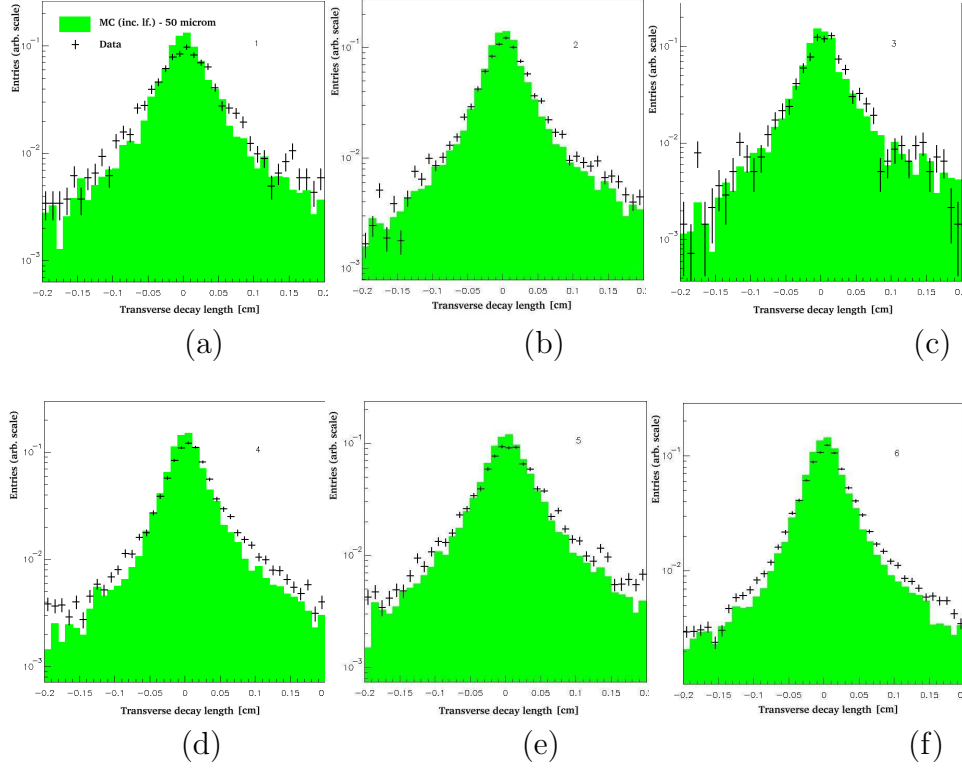


Figure 12.9: MC transverse decay length distributions L_{xy} , smeared by $50 \mu m$, compared to the data in bins of azimuthal angle. Figures (a), (c) and (e) shows the regions covered only by 2 cylinders of ladders. The variable L_{xy} does not show any dependence on the azimuthal coverages areas. A dependence of the smearing procedure on ϕ is not found either.

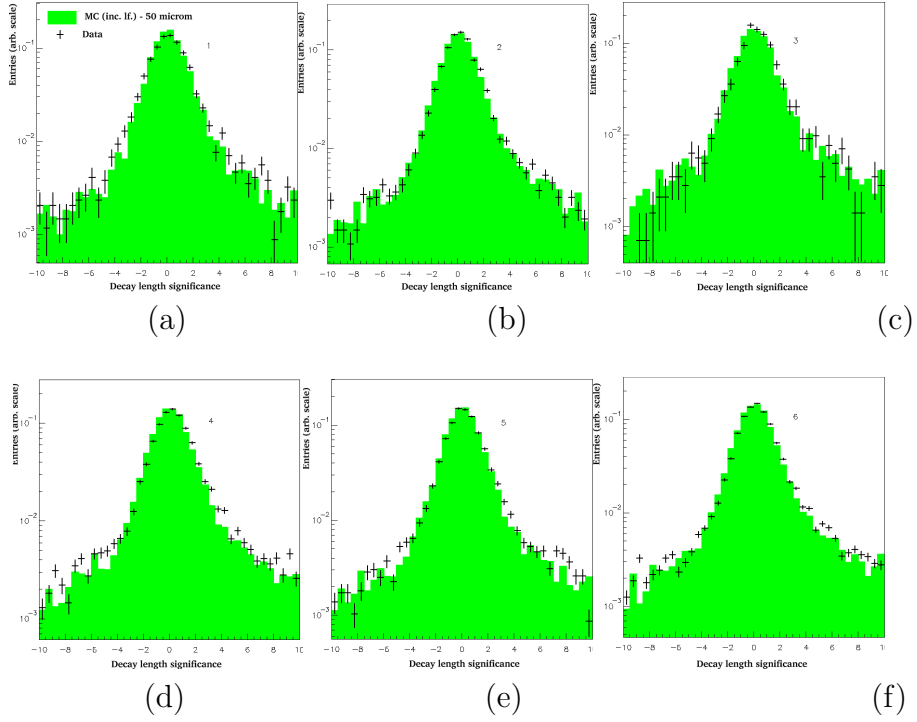


Figure 12.10: MC decay length significance distributions S , smeared by $50 \mu m$, compared to the data in bins of azimuthal angle. Figures (a), (c) and (e) shows the regions covered only by 2 cylinders of ladders. S does not show any dependence on the azimuthal coverages areas.

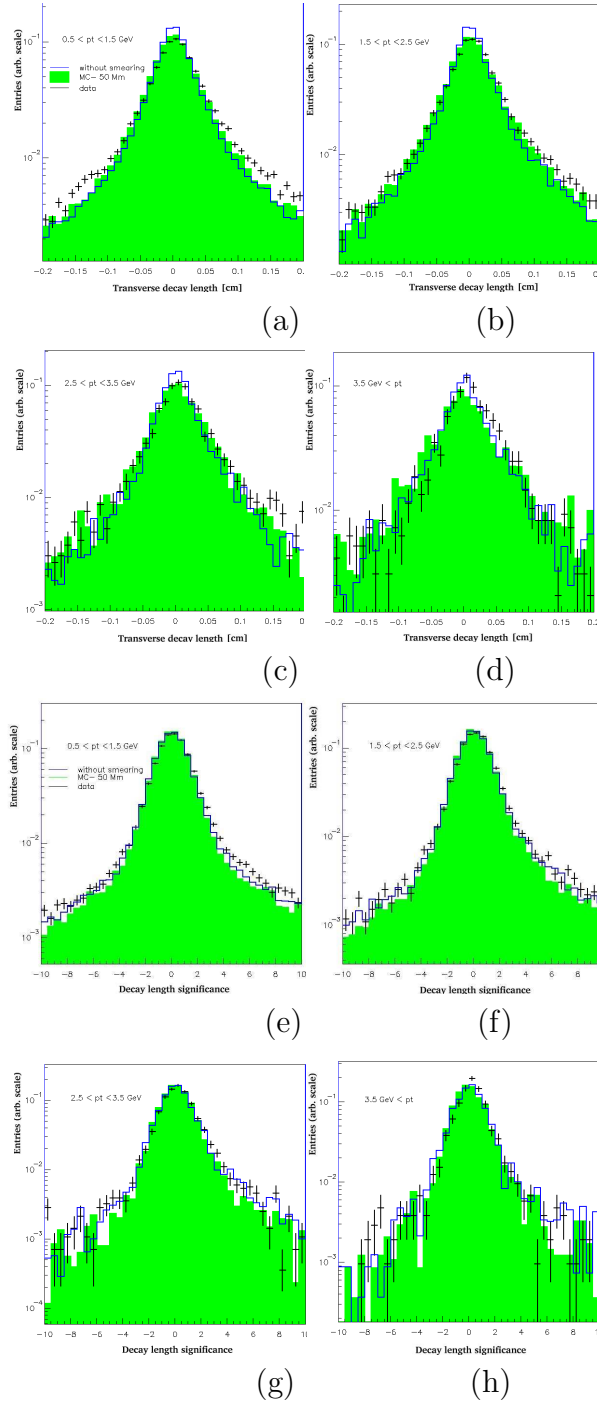


Figure 12.11: MC decay length L_{xy} [(a)-(d)] and significance S [(e)-(h)] distributions, without smearing (solid line) and smeared by $50 \mu m$ (histogram), compared to the data in bins of the average transverse momentum in the secondary vertex. A dependence of the smearing procedure on p_T is not found.

12.6.2 Shift between data and MC

As can be seen from figure 12.6 (before smearing) and 12.7 (after smearing), there is a slight shift of the data to the right with respect to the MC distributions. This shift, also present in the 2004 analysis and seen in other 2005 studies [84], becomes obvious when doing the subtraction ($S^+ - S^-$) (see Fig 12.12 (a)) which, following the same analysis strategy than for the 2004 study (see Sec. 9.1), is the key distribution to obtain the fraction of beauty in the sample. The problem is being intensively investigated. New results

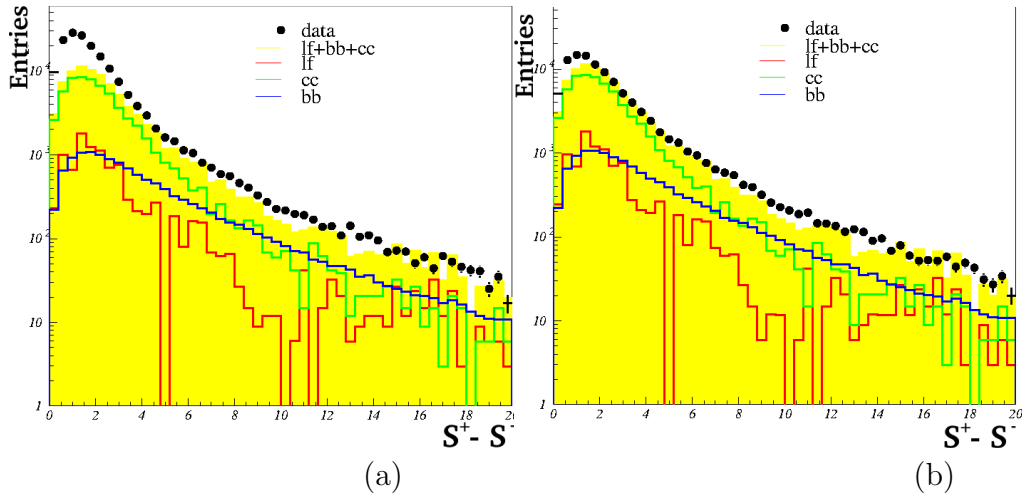


Figure 12.12: Subtracted distribution ($S^+ - S^-$) before redefinition of L_{xy} by 0.002 cm (a) and after (b).

obtained days before the results of this thesis were finished showed the first hints about the source of the shift. The studies point out that the origin of the shift is related with Z_H , the 5th parameter that defines a track in the helix parametrization described in sec. 6.1.2 and which enters in the vertex reconstruction algorithm. Big efforts are being made to understand and fix the problem (see latest results in Appendix B). Other checks performed in order to locate the source of the shift are listed below. The complete set of distributions related to the checks can be found in [84]. Here only the most significant ones are shown:

- Study of the L_{xy} dependence on the azimuthal angle, ϕ , and on the average transverse momentum of the secondary vertices, p_T . Figures 12.9 to 12.11, also used for the MC smearing studies, were reviewed to establish/discard a dependence of the shift on ϕ and/or p_T . The shift is presented in all bins of both variables. No dependence is found. Since

the bins of ϕ were chosen according to the MVD areas covered by 2 or 3 cylinder of ladders and no dependence on them is found, it can be concluded that the problem is not related to the way the tracks are reconstructed in the xy plane. Because the shift does not depend on the p_T either, a relation of it with the track resolution can be as well discarded.

- L_{xy} was calculated with respect to the “reduced” primary vertex instead of the beam spot in order to discard any possible systematic error concerning the beam spot (see Fig. 12.13). The reduced primary vertex is defined as the new primary vertex obtained after removal of the tracks that has been already fitted to a secondary vertex. No relevant differences with the decay length calculated with respect to the beam spot are found.
- L_{xy} was calculated without the projection onto the jet axis, i.e. defined only as the signed distance⁵ between the primary and secondary vertices. The shift is still present so that the source of the shift is independent of the definition of the decay length.
- All the above listed checks were reproduced in a small sample of reprocessed 2006 e^- data. This is the latest available dataset containing new detector alignment and an improved track fitting algorithm [80]. The shift was also found here.

From the above listed checks it could be concluded that the shift does not seem to be related either to the track resolution nor to the tracking procedure. A systematically shifted estimation of the primary vertex or beam spot could also be discarded.

As mentioned above, the problem has now been pinpointed but it still must be understood and fixed. Until the problem is completely solved, a temporary solution is presented within this thesis and will be discussed in the following section.

12.7 Extraction of beauty signal

The inclusive $b\bar{b}$ jet production cross section as well as the differential cross section with respect to the transverse momentum, $d\sigma/dp_T$, and pseudorapidity, $d\sigma/d\eta$, of the b -jet were measured following the same procedure and in

⁵Analogous to the definition of the transverse decay length in the 2004 data analysis

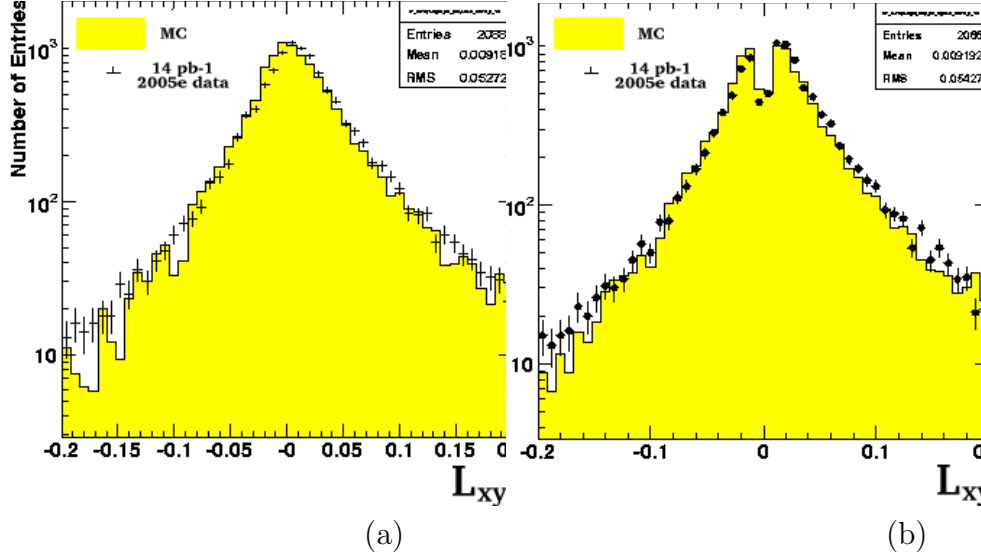


Figure 12.13: (a) Transverse decay length calculated with respect to the reduced primary vertex. (b) Transverse decay length with respect to the beam spot calculated without the projection into the jet axis.

the same kinematic region than in the 2004 analysis and thus, the fit procedure will be only briefly reminded here. In this case though, the decay length significance was used instead of the decay length itself which is, as stated before, one of the greatest improvements of this new analysis. Analogous to the 2004 analysis, in order to reduce the uncertainty due to resolution effects and to cancel most of the light flavour contribution, the contents of the negative bins in the significance distributions were subtracted from the contents of the corresponding positive bins. The subtracted distributions were combined with the secondary vertex mass M_{SV} by calculating $(S^+ - S^-)$ in three different mass regions: $M_{SV} < 1.4 \text{ GeV}/c^2$, $1.4 \leq M_{SV} < 2. \text{ GeV}/c^2$ and $M_{SV} \geq 2. \text{ GeV}/c^2$. Another improvement with respect to the 2004 analysis must be pointed out here. In contrast to the 2004 analysis, the distribution of the secondary vertex mass is now properly described by the MC simulation over the full main range so that further corrections are not needed as before (see Fig. 12.14 (a)).

But, as it was seen in figure 12.12, the subtracted significance distribution shows an important disagreement for low values of $(S^+ - S^-)$, due to the shift between the data and MC. Until the problem is completely solved, a temporary solution has been adopted consisting on the redefinition of the decay length for the data as $L'_{xy} = L_{xy} - 0.002$, where 0.002 is the estimated shift in centimeters. This redefinition will be taken into account within the

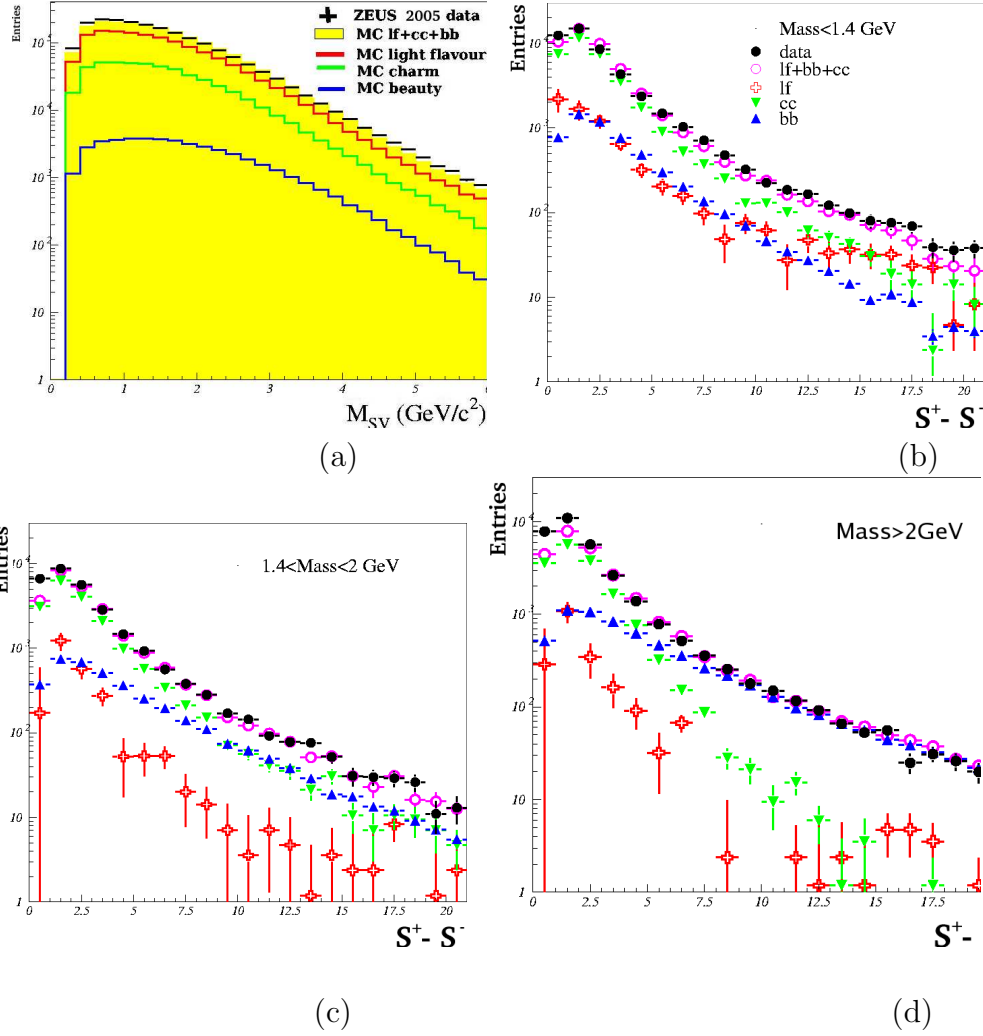


Figure 12.14: (a) Secondary vertex mass distribution of the data compared to the different MC flavour contributions. Figures (b) to (d) $(S^+ - S^-)$ -distributions in bins of M_{SV} . The MC distributions are normalized to the data luminosity.

systematic uncertainties, however one would expect this not to affect much beauty since the shift mostly concerns the region of low $(S^+ - S^-)$ values where charm is the dominant contribution. Figure 12.12 (b) shows the effect of this redefinition on the subtracted distributions. The improvement is remarkable. Figures 12.14 (b)-(c) show the subtracted distributions $(S^+ - S^-)$ in the three bins of M_{SV} .

The fraction of b quarks of the data is extracted using a least squares simultaneous fit to the subtracted distributions $(S^+ - S^-)$ in the second and third bins of M_{SV} where the charm and the beauty contributions dominate respectively. Each MC flavour contribution is then scaled by the obtained factors which give the best fit to the measured $(S^+ - S^-)$ distribution. The statistical errors of the data and MC simulation are considered in the fit. As it is seen in figure 12.14, the first two bins of the subtracted distributions $(S^+ - S^-)$ are still badly described by the MC and therefore are not included into the final fit since they are not expected to be sensitive to beauty in any case. The results of the global fit are depicted in figure 12.15. The χ^2 values shown in all the figures along this chapter are not divided by the number of degrees of freedom which for a fit with two parameters and a total of 10 fitted bins corresponds to $n.d.f = 8$.

12.7.1 Systematic uncertainties

The systematics uncertainties affecting the cross sections are still under evaluation. The full errors are computed as quadratic sum of the statistical and the systematic errors, where the systematic errors are only partially calculated. A list of different systematic checks performed so far and included in all the results and figures is given below. As just stated above, the list is not yet completed. For the differential cross sections the systematic errors are calculated for each bin and are included in all the results and figures.

- **Fit systematics/Redefinition of L_{xy} :** to check the reliability of method of extraction of the beauty signal, all the fits have been redone using the data distributions before the redefinition of the decay length. As expected the fit lead to very high χ^2 values but the resulting scale factors are compatible within errors with the ones obtained previously (see Fig 12.16 and table 12.2). This indicates that the fit method to obtain the beauty scale factors is robust under such variation of the data distribution i.e. the shift hardly affects the region of high $(S^+ - S^-)$ values which is the one dominated by the beauty contribution and thus, though the overall χ^2 becomes drastically worse, the beauty scale factor barely changes since it is determined by the high $(S^+ - S^-)$ region, almost free of background. A summary of

the χ^2/ndf and scale factors obtained from the fit for the total kinematical region as well as in bins of p_T and η is given in table 12.2. The systematic uncertainty coming from this redefinition varies from +3.3% to +13.2%, depending on the kinematic bin in which the differential cross section is being evaluated. For the total cross section, the systematic error is +9.5%.

- **MVD hit efficiency:** In an independent study [65], the ratio of the MVD hit efficiency in data and MC is found to be almost flat at $\sim 95\%$. This study was based on single track calculations. The analysis presented in this thesis deals with secondary vertices and thus, the correlation between tracks must be taken into account. As first estimation, the following procedure was applied to account for the effect of the MVD hit efficiency on the determination of secondary vertices. The acceptance is scaled by 0.95^n , where n is the average number of tracks in a secondary vertex (estimated from Fig. 12.2 (a)). Then, a systematic error of 15% is taken. It should be pointed out again that this is only a gross first estimation and that a dedicated study will follow.

- **Luminosity correction:** The integrated luminosity collected by the ZEUS experiment during the 2005 data taking period is known with an error of $\pm 2.7\%$ [53], which leads to systematic uncertainties of $\pm 2.7\%$.

	redefining L_{xy}		w/o redefining L_{xy}	
	χ^2/ndf	P_b	χ^2/ndf	P_b
Total	7.3/8.	0.95 ± 0.04	97.6/8.	0.86 ± 0.04
$7 \leq p_T < 11 \text{ GeV}$	6.0/8.	0.91 ± 0.06	67.0/8.	0.79 ± 0.06
$11 \leq p_T < 16 \text{ GeV}$	6.2/8.	0.91 ± 0.06	28.9/8.	0.84 ± 0.06
$p_T \geq 16 \text{ GeV}$	16.4/8.	1.21 ± 0.10	10.2/8.	1.17 ± 0.10
$-1.6 \leq \eta < -0.6$	10.9/8	0.65 ± 0.11	31.9/8.	0.59 ± 0.11
$-0.6 \leq \eta < 0.4$	11.5/8	0.81 ± 0.05	69.6/8.	0.74 ± 0.05
$0.4 \leq \eta < 1.4$	19.7/8	1.16 ± 0.06	18.4/8.	1.05 ± 0.06

Table 12.2: χ^2 values and scale factors obtained from the fit before and after the redefinition of the decay length. The fits were performed in the total kinematical area and in bins of p_T and η .

As stated above, only a part of the systematics uncertainties has been calculated. Other systematic checks affecting the cross section measurement are still ongoing. Amongst these checks are the systematics due to energy scale, dijet selection, description of the shape of the secondary vertex mass by the MC simulation and fit related systematics uncertainties such as the implementation of the simultaneous fit of the $(S^+ - S^-)$ in the three mass

regions with three free parameters instead of only fitting the last two bins with two free parameters. The total systematic error calculated so far, is the quadratic sum of the individual changes in the total and differential cross sections. In table 12.3, the total systematic errors for the total cross section are listed.

Description of systematic checks	$\Delta\sigma(\%)$
Fit systematics/Redefinition of L_{xy}	+9.5%
Luminosity measurement	$\pm 2.7\%$
MVD hits efficiency	$\pm 15\%$
Total	+17.9% -15.2%

Table 12.3: Summary of the systematic errors for total cross section measurement.

12.8 First preliminary results

The total measured cross section for inclusive beauty production is:

$$\sigma_{tot}(ep \rightarrow e\bar{b}bX \rightarrow ejjX') = 861. \pm 36. \text{ (stat.) } {}^{+131.}_{-154.} \text{ (syst.) } pb. \quad (12.3)$$

The measurement can be compared to the NLO QCD prediction [59] already given in chapter 10 and reminded below. A good agreement within the still large errors is found:

$$\sigma_{tot}^{NLO} = 836. {}^{+173.}_{-253.} pb. \quad (12.4)$$

The differential cross sections $d\sigma/dp_T$ and $d\sigma/d\eta$ are shown in Table 12.4 and 12.5 and figures 12.19 (a) and (b). For the $d\sigma/dp_T$ distribution, the same p_T binning was chosen than for the 2004 analysis in order to compare the measurements with the ZEUS published result and with the NLO QCD predictions. All the results are compatible with the absolute predictions of the MC model PYTHIA and the NLO QCD calculations corrected for hadronization effects.

Fig. 12.20 shows a comparison between the differential cross sections obtained in the 2004 data analysis and the 2005 data analysis. The full error bars are the quadratic sum of the statistical and the systematic errors, where the systematic errors are only partially calculated for the 2005 data analysis. While the comparison of the $d\sigma/dp_T^{b-jet}$ distribution shows a good agreement between both analysis, the $d\sigma/d\eta^{b-jet}$ distribution shows discrepancies between both results. This can be in part due to the missing systematic errors

of the 2005 results. The fit procedure is also more sensitive to variations in bins of η , being quite robust in the measurements as a function of p_T , leading to higher systematic uncertainties in the calculation of $d\sigma/d\eta^{b-jet}$. These disagreements are still under investigation.

p_T^{b-jet}	$d\sigma/dp_T^{b-jet} \pm stat. \pm syst.$	published ZEUS result
$7 \leq p_T < 11 \text{ GeV}$	$125.3 \pm 8.3^{+25.2}_{-19.0}$	$137. \pm 14^{+21}_{-27}$
$11 \leq p_T < 16 \text{ GeV}$	$30.8 \pm 2.0^{+5.3}_{-4.7}$	$43.8 \pm 5.5^{+7.7}_{-12.0}$
$p_T \geq 16 \text{ GeV}$	$4.2 \pm 0.3^{+0.6}_{-0.6}$	$5.7 \pm 1.2^{+1.03}_{-0.9}$

Table 12.4: Differential cross section for b-jets as a function of p_T^{b-jet} . For comparison, the results are shown together with the previous ZEUS published results [33].

η^{b-jet}	$d\sigma/d\eta^{b-jet} \pm stat. \pm syst.$	published ZEUS result
$-1.6 \leq \eta < -0.6$	$101.0 \pm 17.1^{+17.9}_{-15.3}$	$152 \pm 29^{+24}_{-31}$
$-0.6 \leq \eta < 0.4$	$281.6 \pm 17.4^{+49.3}_{-42.8}$	$356 \pm 41^{+59}_{-53}$
$0.4 \leq \eta < 1.4$	$352.6 \pm 18.2^{+63.1}_{-53.6}$	$275 \pm 45^{+53}_{-73}$

Table 12.5: Differential cross section for b-jets as a function of η^{b-jet} . For comparison, the results are shown together with the previous ZEUS published results [33].

12.9 Outlook

Profiting from higher statistics than in any other previous beauty analysis performed in ZEUS so far, the measurement of the differential cross section $d\sigma/dp_T$ was repeated extending for first time the measurement to higher p_T values. The NLO QCD predictions for the new kinematical range need still to be done, therefore the results are only compared to the absolute PYTHIA prediction (see Fig. 12.21)

	χ^2/ndf	P_b	$d\sigma/dp_T^{b-jet} \pm stat.$
$7 \leq p_T < 11 \text{ GeV}$	6.0/8.	0.91 ± 0.06	125.3 ± 8.3
$11 \leq p_T < 16 \text{ GeV}$	6.2/8.	0.91 ± 0.06	30.8 ± 2.0
$16 \leq p_T < 21 \text{ GeV}$	9.2/8.	0.27 ± 0.11	9.1 ± 0.8
$21 \leq p_T < 26 \text{ GeV}$	12.3/8.	0.63 ± 0.19	1.2 ± 0.4
$p_T > 26 \text{ GeV}$	6.7/8.	1.1 ± 0.81	0.4 ± 0.3

Table 12.6: χ^2 values and scale factors obtained from the fit before and after the redefinition of the decay length extended to higher p_T values.

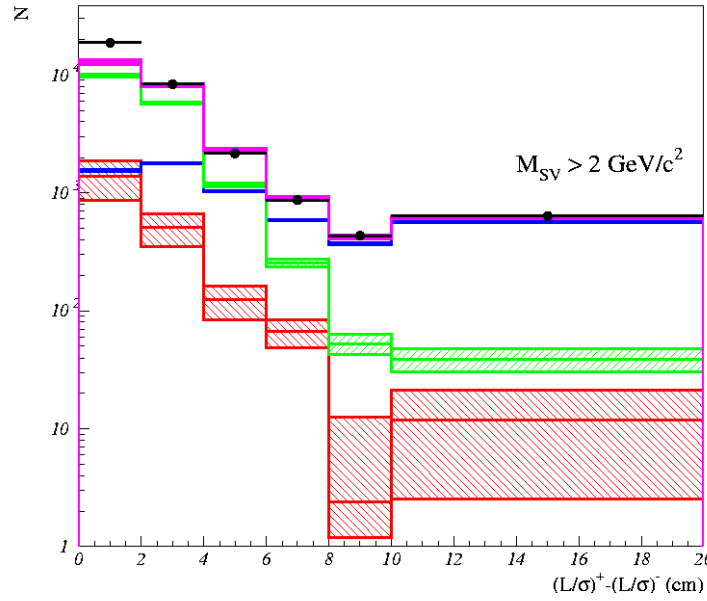
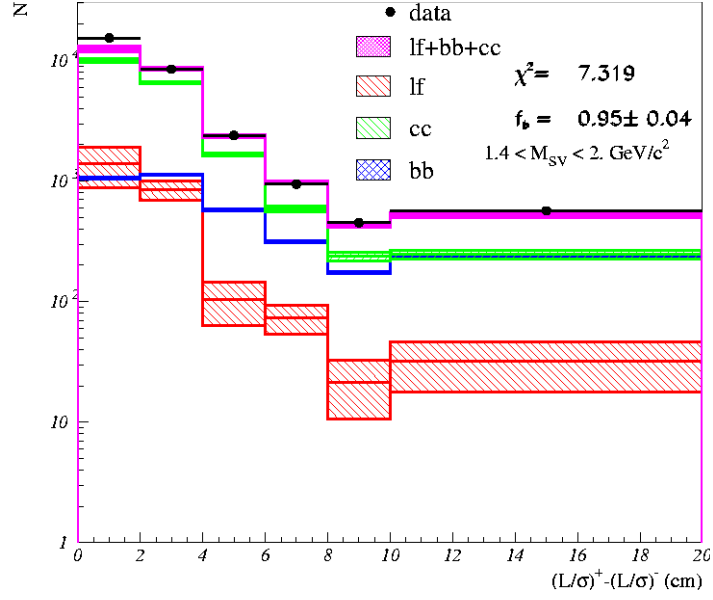


Figure 12.15: Results of global 2-parameters simultaneous fit of $(S^+ - S^-)$ for the bins: (top) $1.4 \text{ GeV}/c^2 < M_{SV} < 2. \text{ GeV}/c^2$ in which the charm contribution dominates and (bottom) $M_{SV} > 2. \text{ GeV}/c^2$ dominated by beauty at large $(S^+ - S^-)$ values.

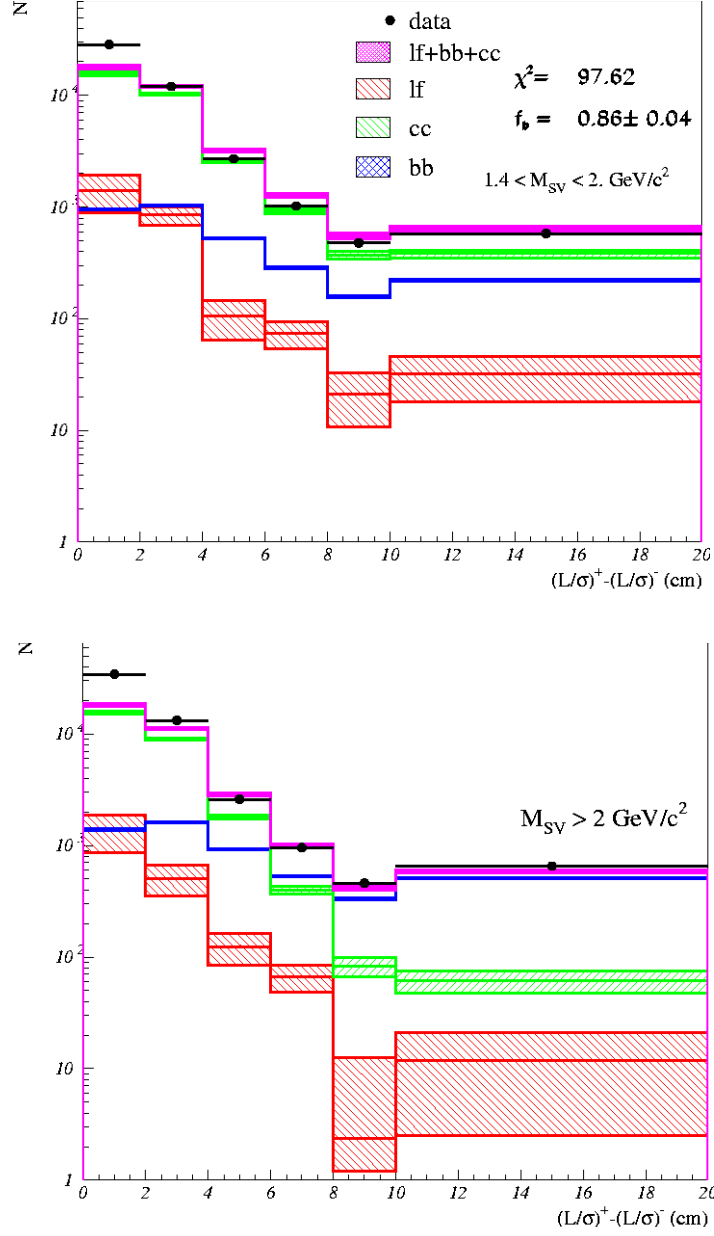


Figure 12.16: Results of global 2-parameters simultaneous fit, before redefining the decay length, of $(S^+ - S^-)$ for the bins: (top) $1.4 \text{ GeV}/c^2 < M_{SV} < 2. \text{ GeV}/c^2$ in which the charm contribution dominates and (bottom) $M_{SV} > 2. \text{ GeV}/c^2$ characterized by beauty.

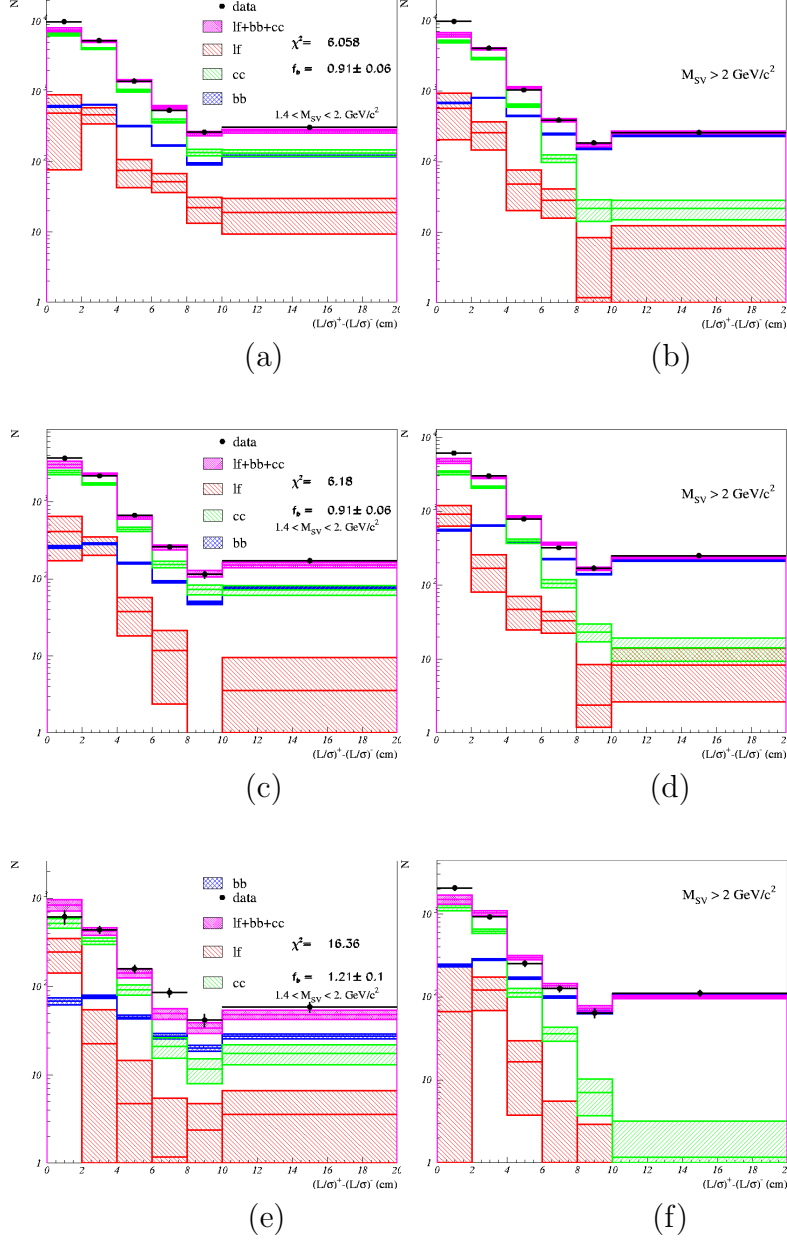


Figure 12.17: $(S^+ - S^-)$ distributions in bins of p_T after the fit to the data points. The bins are $7 \leq p_T < 11$ GeV for plots (a)-(b), $11 \leq p_T < 16$ GeV for plots (c)-(d) and $p_T \geq 16$ GeV for plots (e)-(f). The plots on the left side correspond to the distributions in the second bin of the secondary vertex mass, the plots on the right side to the third bin of the secondary vertex mass.

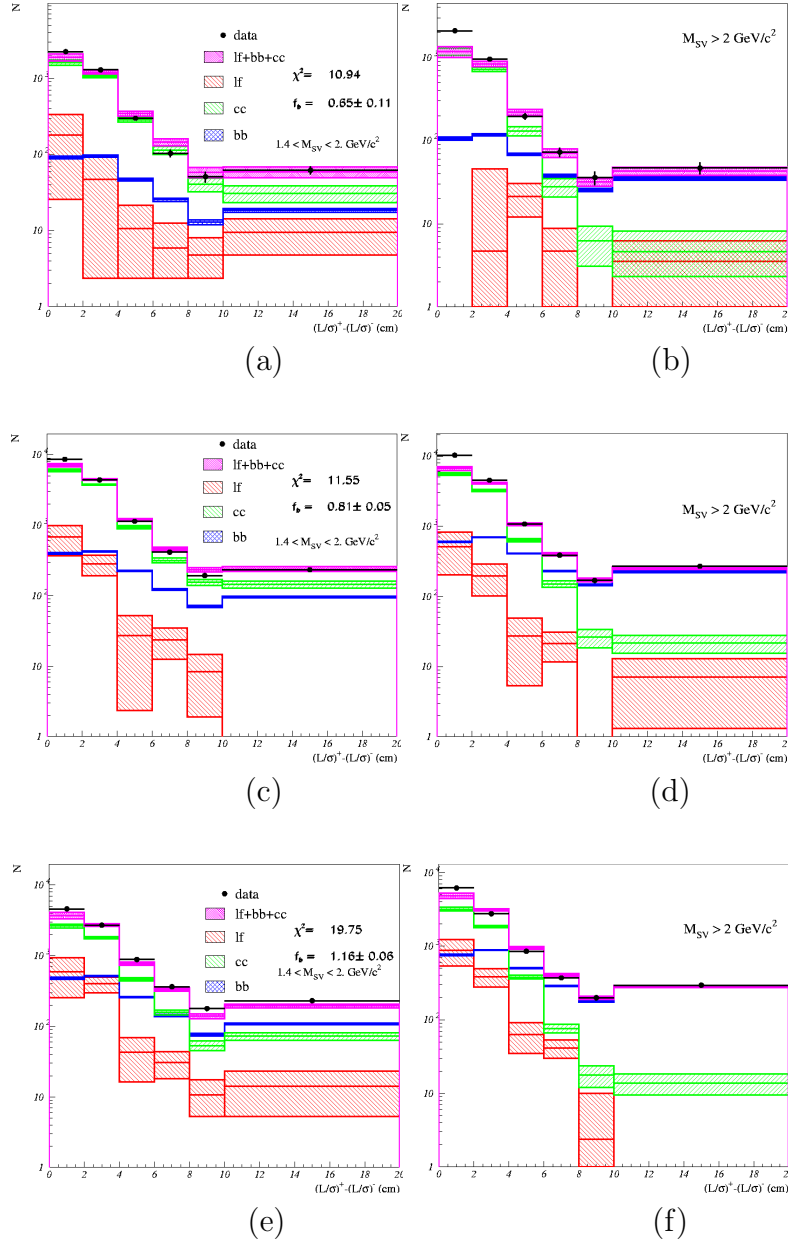


Figure 12.18: $(S^+ - S^-)$ distributions in bins of η after the fit to the data points. The bins are $-1.6 \leq \eta < -0.6$ for plots (a)-(b), $-0.6 \leq \eta < 0.4$ GeV for plots (c)-(d), $0.4 \leq \eta < 1.4$ GeV for plots (e)-(f) and $1.4 \leq \eta < 2.5$ GeV for plots (g)-(h). The plots on the left side correspond to the distributions in the second bin of the secondary vertex mass, the plots on the right side to the third bin of the secondary vertex mass.

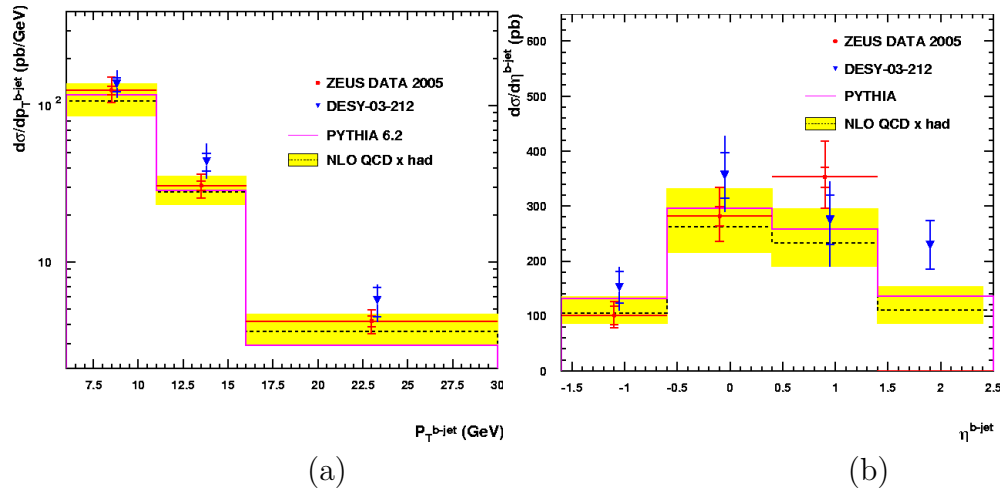


Figure 12.19: Differential cross sections as a function of (a) the transverse momentum p_T^{b-jet} and (b) the pseudorapidity η^{b-jet} of the jet obtained with the 2005 dataset. The data (red circles) are compared to the predictions of PYTHIA (solid line) and to a previous ZEUS measurement [33] (blue triangles moved slightly to the right for a better comparison, see text). The full error bars are the quadratic sum of statistical and systematic uncertainties (inner and outer bars respectively). The data are also compared to the predictions of NLO QCD (dashed line histogram) corrected to hadron level. The band around the NLO prediction represents the uncertainty on the theoretical prediction (see Sec. 2.8.2)

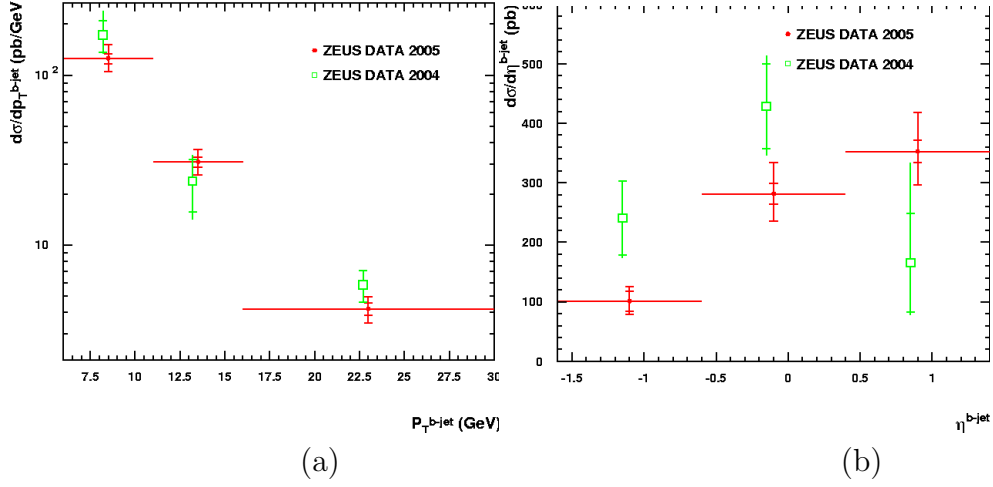


Figure 12.20: Comparison of the differential cross sections as a function of (a) the transverse momentum p_T^{b-jet} and (b) the pseudorapidity η^{b-jet} of the jet obtained with the 2005 dataset with the one obtained with the 2004 dataset. The 2005 results (red circles) are compared to the previous measurement performed with the 2004 dataset (green squares). The full error bars are the quadratic sum of statistical and systematic uncertainties (inner and outer bars respectively). Note that the systematic error of the 2005 analysis is still incomplete.

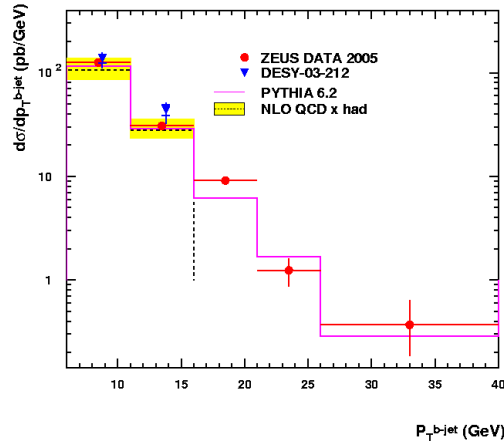


Figure 12.21: Differential cross sections as a function of the transverse momentum p_T^{b-jet} of the jet obtained with the 2005 dataset. The measurement has been extended to higher p_T values. The data (red circles) are compared to the predictions of PYTHIA (solid line). Only statistical errors are depicted. The data are also compared to the predictions of NLO QCD (dashed line histogram) corrected to hadron level. The band around the NLO prediction represents the uncertainty on the theoretical prediction (see Sec. 2.8.2)

12.10 Conclusions

After the promising results obtained in a first exploratory analysis with the 2004 data, a new more precise study of the inclusive beauty photoproduction is ongoing with the 2005 dataset. This analysis is also motivated by improved tracking and vertexing techniques which allow, through a combination of the characteristic distributions of the decay length significance and the invariant mass of the secondary vertices, to exploit the full micro vertex detector potential to extract directly the beauty content in the sample using inclusive dijets. The feasibility of the tagging method developed in this work has been confirmed through first measurements of inclusive beauty photoproduction in dijet events.

This tagging method is intended to be used as well for charm measurements as soon as the systematics are completely understood. As it was seen, the main systematics affect the kinematic regions in which the charm contribution is dominant, making the determination of the charm fraction not yet reliable.

As a final conclusion from this new analysis with 2005 data, the b -quark differential cross section as a function of the quark transverse momentum, $d\sigma(ep \rightarrow bX)/dp_T^b$, is shown in figure 12.22. It was obtained extrapolating the cross section for dijet events using the NLO QCD prediction corrected for hadronization. The \bar{b} quark was not considered in the definition of the b -quark cross section. This extrapolation allows the direct comparison not only with previous ZEUS measurements given at the b -quark level but also with H1 results [32]. All measurements are consistent and in agreement with the NLO QCD predictions.

Moreover, thanks to both the 4 times higher statistics and the better reconstruction from which the study profits, the measurement of the differential cross section $d\sigma/dp_T^{b-jet}$ could be for first time extended towards higher p_T values. The results presented in this thesis aim to contribute to the clarification of the remaining differences observed between theory and experiment for the beauty quark production.

HERA

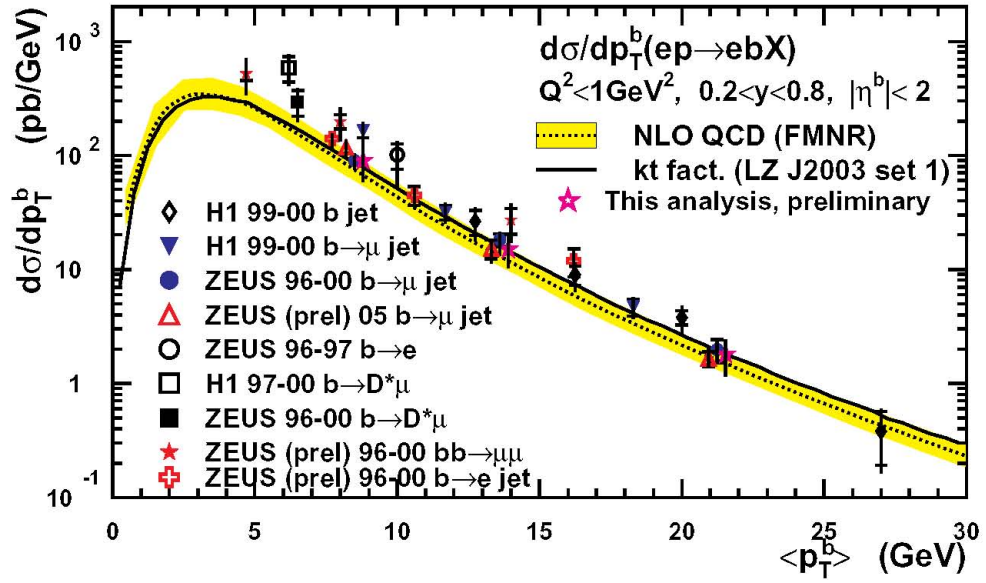


Figure 12.22: Differential cross sections for b -quark as a function of the b -quark transverse momentum p_T^b for b -quark pseudorapidity $|\eta^b| < 2$ and for $Q^2 < 1 \text{ GeV}$ and $0.2 < y < 0.8$. The pink crosses are the results obtained in this analysis with the ZEUS 2005 data. All published results from the ZEUS and H1 experiment are shown [32], including new preliminary results. The dashed line is the NLO QCD prediction with its theoretical uncertainty (shaded band).

Appendix A

Detection principle of silicon detectors

Semiconductor silicon detectors are used widely in high-energy physics. They have many advantages with respect to other solid-state detectors which made of them the best choice when it came to decide by the ZEUS collaboration which kind of vertex detector would best satisfy the physics requirements.

While the ZEUS micro vertex detector was described in detail in Chapter 4, here an introduction to the detection principle of semiconductor detectors is given. The choice of silicon as the semiconductor material will be justified as well. A more complete description of semiconductor detectors can be found e.g. in [81].

A.1 Detection principle

If a charged particle travels through matter, the particle loses energy as a result of different processes (e.g. through inelastic scattering from atomic electrons). The energy loss per unit length of a particle is described by the Bethe-Bloch equation:

$$-\frac{dE}{dx} = 2\pi N_A m_e r_e^2 c^2 \rho \frac{z^2}{\beta^2} \frac{Z}{A} \left[\ln \left(\frac{2m_e \gamma^2 v^2 W_{max}}{I^2} \right) - 2\beta^2 \delta - 2\frac{C}{Z} \right] \quad (\text{A.1})$$

where N_A is Avogadro's number, m_e the electron mass, r_e the classic electron radius, ρ the density of the traversed material, z the charge of the incident particle, $\beta = v/c$ the velocity of the particle in units of speed of light, Z and A the atomic number and the atomic mass (in g/mol) of the material, I the effective ionization potential, $\gamma = 1/\sqrt{(1 - \beta^2)}$, W_{max} the maximum energy

transfer in a single collision, δ the density correction and C a shell correction.

The energy loss of the incident particle is used to ionize the semiconductor atoms producing pairs of electron and holes along its track through the detector (see Fig. A.1). As it will be seen in detail in the next sections, the electrons and holes (charge carriers) are separated by an externally applied electric field before they could recombine. The charges are recollected on the surface electrodes located to both sides of the detector. This process produces a current pulse whose integral is a measure for the energy deposited by the particle. The electrode pulse can be detected using charge sensitive amplifiers. Through the division of one or both sides of the detector into small strips, a sensitivity to the hit position is obtained allowing the reconstruction of the particle's track.

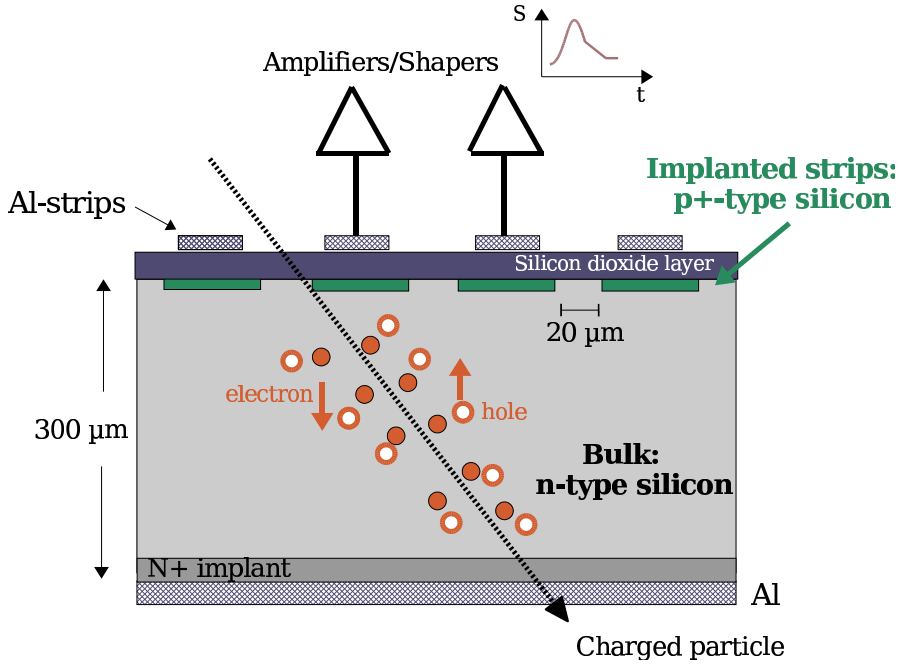


Figure A.1: Schematic cross section of a silicon detector. When an ionizing particle crosses the silicon bulk, many charge carriers (electrons and holes) are produced per unit traversed length. The generated carriers move within the electric field toward the nearest electrode, where they are collected as a current pulse.

A.2 Physical properties of semiconductors

The physical properties of semiconductors can be well described by the energy band model (see Fig. A.2). For crystalline solids, the energy levels of the electrons are structured into bands. The valence band is the uppermost allowed energy band that is completely occupied by electrons at absolute zero ($T = 0K$). On the other hand, the conduction band is the lowermost energy band that is unoccupied by electrons at $0K$ i.e. the upper band of allowed states. The valence band and the conduction band are separated by a band gap in which all energy states are forbidden. The size of this band gap determines whether the material is an insulator ($> 5 eV$) or a conductor (no band gap). For small band gap ($\sim 1 eV$), the material is classified as semiconductor. For semiconductors, at $0K$, all the electrons in the valence band participate in covalent bonding between the lattice atoms but at finite temperature (e.g. room temperature $kT \sim 25 meV$) the electrons can be thermally excited from the valence band to the conduction band leading to creation of electron-hole pairs (conductivity). The properties of a semicon-

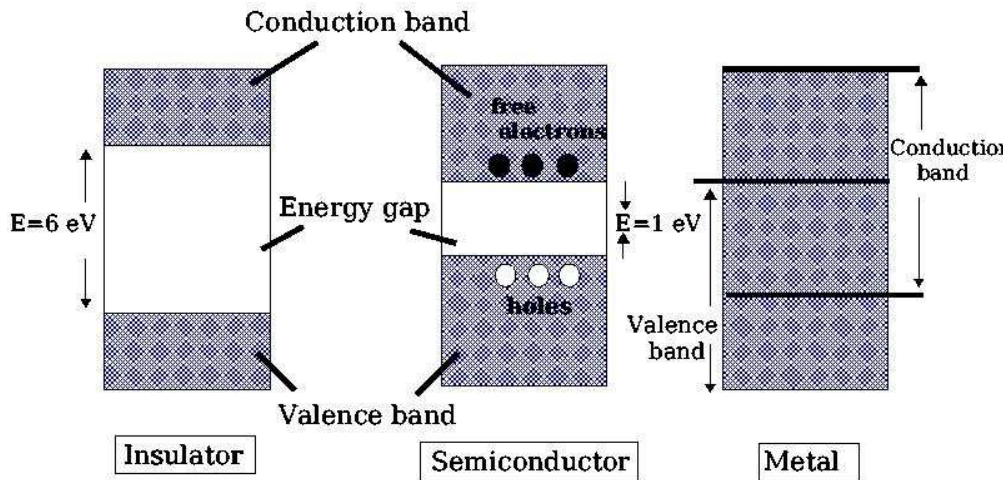


Figure A.2: Energy band structure of conductors, insulator and semiconductors. The valence band is the highest energy band completely filled with electrons at absolute zero. The conduction band is the next highest possible energy band above the valence band. In case of insulators and semiconductors, both bands are separated by a forbidden band gap.

ductor can be modified by adding impurities, donors and/or acceptors, in a process called *doping*. Donors are atoms with 5 electrons¹ in the outer shell

¹taking silicon as semiconductor material

from which only 4 are used for the covalent bonding with the neighboring atoms while the fifth is donated to the material leading to an excess of negative charge carriers. Acceptors have only 3 electrons in the outer shell so that to complete the covalent bonds to the lattice atoms, it captures electrons from the valence band of the material i.e. produces holes. According to the band model, adding a donor or an acceptor to the crystal is equivalent to the creation of localized energy levels within the band gap closer to the conduction or valence band respectively. Semiconductors doped with donor impurities are called *n-type* while those doped with acceptors are known as *p-type*.

A.3 $p - n$ diode junction

The operation of semiconductors as detectors is usually directly related to the use of $p - n$ diode junctions. The reason is to suppress the intrinsic noise of the detector as much as possible i.e. to increase the signal-to-background ratio. When a charged particle crosses a semiconductor material, it creates electron-hole pairs. These electrons and holes recombine after a short time leading to a loss of the information about the traversing particle. Therefore, an external voltage is applied to the semiconductor in order to separate the charge carriers and keep the information. However, generation of electron-hole pairs is also possible by simple thermal excitation. In this case, the consequent recombination of the electrons and holes leads to an intrinsic current (noise) which, depending on the resistivity of the semiconductor material, can be of the same order than the signal generated by the traversing charged particle. If it happens, the semiconductor is clearly not suitable for particle detection. As just stated above, a $p - n$ junction is used to improve the signal-to-background ratio.

A $p - n$ junction consists of putting in contact a region of n-type semiconductor material with a region of p-type material. The n-type region is characterized by many free electrons in the conduction band and almost no holes in the valence band, while in the p-type region there are many free holes in the valence band and almost no electrons in the conduction band. Thermal excitation leads to a diffusion of the electrons from the n-side and the holes from the p-side to the adjacent region of the other side respectively. The different concentrations of holes and electrons participating in the diffusion creates an excess of positive charge in the n-region and an excess of negative charge in the p-region. As a consequence of it, an electric field gradient across the junction is built up. The contact region is called *depletion* zone and is characterized by the reduced amount of free charge carriers in

it (thanks to the electric field). Any charge created in the depletion region will be collected by the electric field and can be detected by electronics connected on both sides of the junction. Charges originated outside this region will simply recombine until thermal equilibrium is reached, and thus remain undetected. Therefore, the aim is to have a large depletion zone which means more volume sensitive to detection. This can be achieved by making the $p-n$ junction to work as a diode i.e. modifying the potential difference between the two sides of the junction. The applicable voltage is limited by the resistance of the semiconductor since after some point the junction breaks down and it becomes conductive.

A.4 Why silicon?

Silicon is chosen commonly as semiconductor material since it has properties that make it particularly suitable for the detection of ionizing radiation. Some of the most important physical properties of silicon and their direct connection to the detector's characteristics are listed below:

- **small band gap (1.12 eV at 300K):** i.e. the average energy required to generate an electron-hole pair is low ($\approx 3.5 \text{ eV}^2$). This leads to a large number of charge carriers per unit energy loss of ionizing particles which is translated into a high intrinsic energy resolution
- **low thermal energy ($\approx 1/40 \text{ eV}$):** i.e. little cooling is required
- **high density (2.33 g/cm^3):** leads to a large energy loss per traversed material length of ionizing particles. It allows the use of thin detectors
- **high carrier mobility ($\mu_e = 1450 \text{ cm}^2/Vs$, $\mu_h = 450 \text{ cm}^2/Vs$):** i.e. the charge carriers move very fast through the silicon lattice which leads to a rapid charge collection of the order of $\sim 10 \text{ ns}$
- **precise reconstruction of the position is possible:** for position sensitive silicon sensor in which one or both sides of the diode can be segmented into strips of the order of $\sim 10 \text{ }\mu s$.

Summarizing, silicon detectors are widely applied in high energy physics for their well known characteristics. Specially, their unmatched energy and spatial resolution as well as their excellent response time make of them advantageous over the traditional gas (or liquid) filled detectors.

²For comparisons: the average energy required to create an electron-hole pair in Argon gas is $\approx 15 \text{ eV}$

Appendix B

Latest results on the description of L_{xy} by MC

The distributions of the decay length and decay length significance are not correctly described by the MC simulation. Moreover, there is a systematic shift to the right of the data shape with respect to the MC. This problem as well as a whole list of checks performed in order to find the source of the shift was described in detail in the section 12.6.2. Few days before this thesis was completed, new results finally showed the first hints about where this shift is coming from. Big efforts are being made in order to understand and consequently fix the problem. In this appendix, a summary of the ongoing studies is shown. Updates are coming in a daily basis and can be found in [84].

B.1 Dependence on Z_H and θ_H

It was seen in section 6.1.2 that any trajectory of a particle can be parametrized as an axial helix.

This helix is defined through 5 parameters $\vec{p} = (\phi_H, Q/R, QD_H, Z_H, \cot \theta_H)$ and its corresponding covariance matrix. Even if the transverse decay length is calculated in two dimensions from the position of the primary and secondary vertex in the XY - plane, the vertex itself is a three dimensional object and consequently its reconstruction was, so far, performed in three dimensions. However, the studies show that the shift seem to be related to the Z coordinate of this parametrization. In fact, the calculation of the decay length in bins of the Z coordinate and of the polar angle θ shows a clear dependence of the decay length on these two variables. Figure B.1 show the variation of the shift as a function of Z and θ , where θ was divided in

steps of 20° and Z in steps of 12 cm respectively. Moreover, if the tracks parameters that enter the vertexing algorithm are modified by setting the coordinates Z_H and $\cot\theta_H$ to zero as well as all covariance matrix elements depending on them, the shift disappears and the agreement between data and MC simulations becomes very good (see Figure B.2).

The problem is now located but still more work must be done in order to solve it in a satisfactory way so that both, the measurement presented in this thesis as other similar analysis based on decay length distributions, can be done with much higher accuracy, derivated from a proper description of the data by the MC simulations.

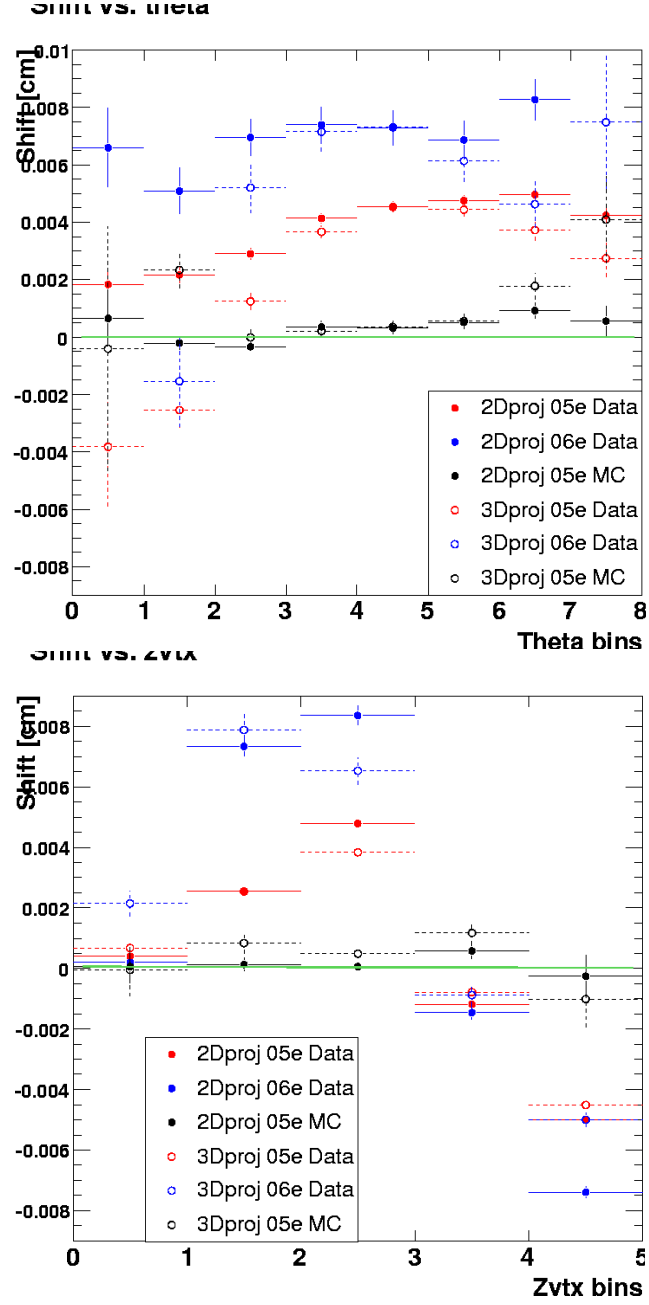


Figure B.1: Dependent of the shift of the decay length distribution on the polar angle θ (top) and the Z coordinate (bottom). Shown are the studies performed for the 2005 and 2006 datasets as well as for the decay length calculated in two and three dimensions.

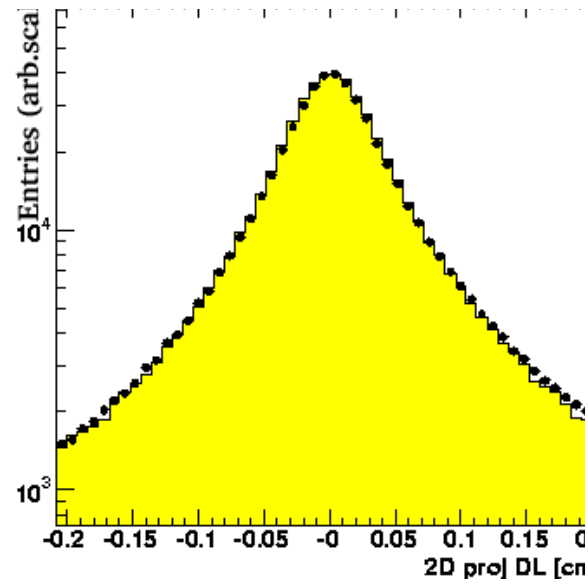


Figure B.2: Transverse decay length distribution after 2D vertexing.

Bibliography

- [1] S. Schlenstedt A. Lopez-Duran. electron finder efficiencies and impurities. a comparison between sinistra95, em and emnet. *ZEUS-NOTE 99-077*, 1999. New version from July 2000.
- [2] I. et al. Abt. Measurement of the b anti-b production cross-section in 920-gev fixed target proton nucleus collisions. *Eur.Phys.J.*, C26:345, 2003.
- [3] C. et al. Albajar. Beauty production at the cern p anti-p collider. *Phy. Lett.*, B256:121, 1991.
- [4] Parisi Altarelli, Guido. Asymptotic freedom in parton language. *Nucl.Phys.*, B126:298, 1977.
- [5] Amadeus. Amadeus, 2007. URL http://www-zeus.desy.de/ZEUS_ONLY/analysis/Amadeus/.
- [6] J. et al. Andruszkow. Luminosity measurement in the zeus experiment. *Acta Phys.Polon.*, B32:2025, 2001.
- [7] B.Andersson. *The Lund Model*. Cambridge Monographs on Particle Physics, 1983.
- [8] Gennady M. Briskin. Diffractive dissociation in ep deep inelastic scattering. *DESY-THESIS-98-036*, 1998.
- [9] Zerwas C. Peterson D.Schlatter, Schmitt. Scaling violations in inclusive e+ e- annihilation spectra. *Phys.REv*, D27:105, 1983.
- [10] Yuri L. Webber Catani, Dokshitzer. The k-perpendicular clustering algorithm for jets in deep inelastic scattering and hadron collisions. *Phys.Lett.*, B285:291, 1992.
- [11] G.Sterman C.Collins, D.Soper. Soft gluons and factorization. *Nucl.Phys.*, B308:833, 1988.

- [12] The ZEUS Coll. Evtake, 2007. URL http://www-zeus.desy.de/~ifhuta/ZEUS_ONLY/evtake/.
- [13] The ZEUS Coll. A microvertex detector for zeus. *DESY-PRC 97-01, ZEUS-NOTE-97-006*, 1997.
- [14] H1 Collanoration. Measurement of charm and beauty dijet cross sections in photoproduction at hera using the h1 vertex detector. *DESY-03-212*, 2003. Submitted to Eur.Phys.J.C.
- [15] M. et al. Derrick. Design and construction of the zeus barrel calorimeter. *Nucl.Instrum.Meth*, A309:77, 1991.
- [16] E.J.Williams. Nature of the high-energy particles of penetrating radiation and status of ionization and radiation formulae. *Phys.Rev*, 45: 729–730, 1934.
- [17] E.Maddox. Study of heavy quark production at hera using the zeus microvertex detector. Master’s thesis, University of Amsterdam, NIKHEF, 2004.
- [18] B.Foster et al. The design and construction of the zeus central tracking detector. *Nucl.Instr. and Meth.*, A338:254, 1994.
- [19] Bloom et al. High-energy inelastic ep scattering at 6-degrees and 10-degrees. *Phy.Rev.Lett.*, 23:23, 1969.
- [20] D.Acosta et al. Measurement of the j/ψ meson and b-hadron production cross sections in p anti-p collisions at $s^{**}(1/2) = 1960\text{-gev}$. *Phy.Rev*, D71:032001, 2005.
- [21] D.Acosta et al. Measurement of the b^+ total cross-section and b^+ differential cross-section $d\sigma/dp(t)$ in p anti-p collisions at $s^{**}(1/2) = 1.8\text{-tev}$. *Phy.Rev*, D65:052005, 2002.
- [22] D.Acosta et al. Measurement of the b^+ total cross section and the b^+ differential cross section $d\sigma/dp_T$ in $p\bar{p}$ collions at $\sqrt{s} = 1.8\text{ tev}$. *Phys.Rev.*, D65:052005, 2002.
- [23] F.Aber et al. Measurement of correlated mu- anti-b jet cross sections in $p\bar{p}$ collions at $\sqrt{s} = 1.8\text{ tev}$. *Phys.Rev.Lett.*, 53:1051, 1996.
- [24] H.L.Lai et al. Global qcd analysis of parton structure of the nucleon: Cteq5 parton distributions. *Eur.Phys.*, J.C12:375–392, 2000.

- [25] I. Bloch et al. A radiation monitor for the zeus detector at hera. *IEEE Transactions on Nuclear Science*, 51:4, 2004. URL <http://www-zeus.desy.de/components/mvradmon/>.
- [26] J.Huth et al. Proceedings of the pdf summer study on high energy physics. In *Proceedings of the PDF Summer Study on High Energy Physics*, Singapore World Scientifics, 1992.
- [27] P.Bordalo et al. Open beauty production in high-energy pi- - tungsten interactions. *Z.Phys.*, C39:7, 1988.
- [28] R. Hall-Willton et al. The ctd tracking resolution. *ZEUS-NOTE-99-024*, 1999. unpublished.
- [29] R.Jesik et al. Bottom production in pi- - be collisions at 515-gev/c. *Phys.Rev.Lett.*, 74:495, 1995.
- [30] S.Abachi et al. Inclusive mu and b quark production cross-sections in $p\bar{p}$ collions at $\sqrt{s} = 1.8$ tev. *Phy.Rev.Lett.*, D70:3548, 1995.
- [31] S.Abbott et al. The $b\bar{b}$ production cross section and angluar correlations in $p\bar{p}$ collions at $\sqrt{s} = 1.8$ tev. *Phy.Lett.*, 487:264, 2000.
- [32] S.Chekanov et al. Beauty photoproduction measured using decays into muons in dijet events in e p collisions at $s^{1/2} = 318$ GeV. *DESY-03-212*, page 32pp, 2003.
- [33] S.Chekanov et al. Beauty photoproduction measured using decays into muons in dijet events in e p collisions at $s^{1/2} = 318$ GeV. *Phy.Rev.*, D70:012008, 2004.
- [34] S.Eidelman et al. Review of particles physics. *Phy.Lett.*, B592, 2005. URL <http://pdg.lbl.gov/>. Particle Data Group.
- [35] T.H.Bauer et al. The hadronic properties of the photon in high-energy interactions. *Rev.Mod.Phys.*, 51:407, 1979.
- [36] T.Sjöstrand et al. *PYTHIA 6.4 PHYSICS AND MANUAL*, 2006.
- [37] V.M.Budnev et al. The two photon particle production mechanism. physical problems. applications. equivalent photon approximation. *Phys.Rep.*, C15:181, 1974.
- [38] W. Fallot-Burghardt et al. Helix user manual v2.1. *HD-ASICS-33-0697*, 1999.

- [39] B.Aubert et al. BABAR Collab. Measurement of the branching fraction for inclusive semileptonic b meson decays. *Phys.Rev.*, D67:031101, 2003.
- [40] fermilab. Fermilab, 2007. URL www.fnal.gov.
- [41] Feynman diagrams. Feynman diagrams, 2007. URL <http://www2.slac.stanford.edu/vvc/theory/feynman.html>.
- [42] F.James. Function minimization and error analysis, 1994. Reference Manual, Version 94.1. Computing and Networks Division CERN.
- [43] G.Hartner. private communication, 2005.
- [44] G. Bishop G.Welch. An introduction to the kalman filter, 2007. URL <http://www.cs.unc.edu/~welch/kalman/>.
- [45] Tobias Haas. private communication, 2005.
- [46] F. Halzen and Martin. *Quark and Leptons*. A John Wiley and Sons, 1984.
- [47] G.F. Hartner. Vctrack briefing: Program and math. *International ZEUS-Note 98-058*, 1998.
- [48] W.Heitler H.Bethe. On the stopping of fast particles and on the creation of positive electrons. *Proc.Roy.Soc.Lond*, A146:83, 1934.
- [49] HERA. A proposal for a large electron-proton collading beam facility at desy. *DESY-HERA-81/10*, 1981.
- [50] hstadie. tlite home page, 2006.
- [51] D.M. et al. Jansen. Measurement of the bottom-quark production cross section in 800 gev/c proton-gold collisions. *Phys.Rev.Lett*, 74:3118, 1995.
- [52] Uta Klein. private communication, 2007.
- [53] Uta Klein and Krystyna Olkiewicz. private communication, November 2007.
- [54] Regis Lefevre. Inclusive and dijet b productions at cdf. *FERMILAB-CONF-05-626-E*, 2005.
- [55] INMOS Limited, editor. *The Transputer Book*. Prentice-Hall, Inc, 2nd edition, 1989.

- [56] H.W.Wyld L.M.Jones. A measurement of beauty quark production cross-section in 800-gev/c proton-gold collisions. *Phys.Rev.*, D17:759, 1978.
- [57] Cacciari M. Qcd predictions for charm and bottom production at rhic. *Phys.Rev.Lett.*, 95:122001, 2005.
- [58] A. Vogt M. Gluck, E. Reya. Photonic parton distributions. *Phys.Rev.*, D46:1973–1979, 1999.
- [59] M.Corradi M. Turcato. private communication, 2006,2007.
- [60] E. Maddox. A kalman filter track fit for the zeus microvertex detector. *ZEUS-03-008*, 2003.
- [61] M.Turcato M.Corradi. Beauty production in dijet events. *ZEUS-NOTE-04-005*, 2004. unpublished.
- [62] M.Erdmann. *The Partonic Structure of the Proton*, volume 138. Springer Tracts in Modern Physics, 1997.
- [63] M.Jechow. Studies of secondary vertex finding for beauty at ep collisions at zeus. Master’s thesis, Humboldt University of Berlin, 2007. Diploma Thesis. Hardcopy at the HU of Berlin.
- [64] M.Seidel. The upgraded interactions regions of hera. *DESY-HERA-00-01*, 2000.
- [65] M.Turcato. private communication, November 2007.
- [66] R. Fruhwirth P.Billoir and M. Regler. Track element merging strategy and vertex fitting in complex modular detectors parametrization of tracks. *Nucl.Instr.and Meth.*, A241:115–131, 1985.
- [67] S.Qian P.Billoir. Fast vertex fitting with a local parametrization of tracks. *Nucl.Instr.and Meth.*, A311:1992, 139.
- [68] R.Brun. Geant 3.13, 1987.
- [69] D.Notz R.Frühwirth M.REgler R.K.Bock, H.Grote. *Data Analysis Techniques for High-Energy Physics*. Cambridge Monographs on Particle Physics Nuclear Physics and Cosmology, 2000.
- [70] R.Mankel. private communication, 2007.
- [71] Tracking Group R.Mankel, S.Spiridinov. private communication, 2007.

- [72] Sakurai. Theory of strong interactions. *Ann.Phys.*, 11:1–48, 1960.
- [73] M.L.Mangano P.Nanson S.Frixione and G.Ridolfi. Heavy quark correlations in photon - hadron collisions. *Nuc.Phys.*, B412:225, 1994.
- [74] P.Palazzi S.M.Fisher. Adamo reference manual for version 3.3, 1993. URL <http://adamo.web.cern.ch/Adamo/refmanual/Document.html>.
- [75] S.Miglioranzi. Beam spot for reprocessed data (2003/2004/2005), 2006. URL http://www-zeus.desy.de/~silvia/ZEUS_ONLY/beamspot.html.
- [76] S.Miglioranzi. Ip smearing, 2006.
- [77] S.Miglioranzi. *Beauty photoproduction at HERA II with the ZEUS Experiment*. PhD thesis, University College London, 2006.
- [78] S.Okubo. Phi meson and unitary symmetry model. *Phys.Lett.*, 5:165, 1963.
- [79] Sasha Spiridinov. New tools and performance of rigorous track fit, 2007. Note on preparation.
- [80] S.Spiridinov. Rigorous track fit, 2007. URL <https://www-zeus-data.desy.de/tracking>.
- [81] S.Tavernier. Silicon detectors, 2007.
- [82] Paolo Bellan (PhD Student). private communication, 2007.
- [83] Verena Schoneberg (PhD Student). private communication, 2007.
- [84] Verena Schoneberg (PhD Student), 2007. URL https://www-zeus.desy.de/zems/ZEUS_ONLY/meeting_show.php#sess2042.
- [85] L.M. Lederman et al. S.W. Herb, D.C. Hom. Observation of a dimuon resonance at 9.5 GeV in 400 GeV proton-nucleus collisions. *Phys.Rev.Lett.*, 39:252, 1977.
- [86] Monica Turcato. Measurement of beauty photoproduction at hera. Master’s thesis, University of Padova, 2004. DESY-THESIS-03-039.
- [87] U.Schneekloth. Hera report, October 2004. talk ZEUS Coll.Meeting Padova.

- [88] U.Schneekloth. The hera luminosity upgrade. *DESY-HERA-98-05*, 1998.
- [89] U.Schneekloth. private communication, 2007.
- [90] V.Roberfroid. Improvement of the kalman filter fit in zeus. *ZEUS-NOTE 06-10*, 2006.
- [91] A. Yagues. Mvd hits smearing home page, 2006.
- [92] U. Holmed ZEUS Coll. The zeus detector. status report 1993, 1993. URL <http://www-zeus.desy.de/bluebook/bluebook.html>. unpublished.

Acknowledgement

Now that this thesis is finally completed, I would like to thank all the people who made it possible in one or other way. I only hope not to forget anybody and if I do it, I hope he or she can forgive me. First of all I would like to thank Nikolaj Pavel, wherever he is, for giving me the opportunity to join the ZEUS collaboration and who always encouraged me since the beginning. With his great energy and enthusiasm he proved himself to be an excellent physicist, professor, supervisor and friend. His fighting spirit and his unforgettable laugh accompanied him until his last days and will accompany me always.

I would like to thank Prof.Dr.Thomas Lohse and Prof.Dr.Robert Klanner, for overtaking the supervision of my work and the helpful advices every time I needed them.

I thank very especially my supervisor “in practice” Achim Geiser. I enjoyed all the fruitful discussions (sometimes up to very late in the night). Thanks for all his new ideas, his support and advice at any time. Without his help this work would have never seen a happy end.

I am very grateful to Massimo Corradi who helped me with the choice of the topic for my PhD. He and Silvia Miglioranza provided the starting point for this work. Their support at this early and very important stage of my thesis and their willing and patience to answer all my question have been priceless.

I also want to thank very much Gerd Hartner, Rainer Mankel and Sasha Spiridinov for their help in all topics related with tracking and vertexing. Their support is for me and in general for the Heavy Flavour Group VERY appreciated.

Sincere thanks to my technical task supervisors, Anna Mastroberardino, James Ferrando and Franco Brasolin. Without their help and unlimited knowledge on the funnel system, no job would have run successfully under my term of office as funnel operator.

I have to make a special mention of all my colleagues of the Heavy Flavour Group, even more especially the *beauty subgroup*. The work atmosphere within this group, and ZEUS in general, has always been great. Lively, supportive and demanding at the same time. It provides the best source of physic discussions that led to new ideas and also motivation and encouraging when needed. Special thanks to the beauty team members Verena Schöneberg, Monica Turcato, Paolo Bellan, Benjamin Kahle...who worked hard directly or indirectly on the many problems and technicalities which led to the production of this analysis. Thanks again!

I also want to thank all the people I met at the many conferences I had the pleasure to attend in these years. Their many feedbacks and interesting ideas were always helpful.

Of course, now I want to give the best thanks to all the people who make possible that these last years were not only great from the professional point of view but also unforgettable within my private life. People who I shared tears and laughs with and made of DESY and Hamburg (the eternal raining city) a fun place to live in.

They became much more than colleagues during all this time and I hope to enjoying their friendship for a lot of years more. Thanks to John Loizides for being always so helpful and nice to me (even if you are still scared of my cat Quillo). Tom Danielson for the uncountable hours of chat talking, for helping me skiing down the most scaring pistes...and don't! I do not thank him for the many jokes I had to listen *smile*. Jerome Whyte for the many conversations and laughs we shared (keywords: Amsterdam, bad monkeys, Tom Cruise alias "Billy Dunne"). He became in a surprisingly short space of time one of my most confident friends.

Thanks in particular to the Spanish crew: Manolo Zambrana for his cooking lessons and for teaching me what patience means (I have to practice it always when he drives ;-P). And the most *infernales gracias* to Elias Ron and Marcos Jimenez (and the lately joined Alejandro Lopez). They bring me to laugh daily with their jokes and their way of life. They also make me think in my free time!! with riddles and discussions about philosophy, religion or any other transcendental topic *arghh* . We may not always agree but we always have a good laugh out of all it.

I want to thank from heart three people who are very special to me: my friends Michel, Pat and Stefan. There are no words to describe how much they mean in my life but they definetely know it.

Last but not less, I want to thank my family, without whom I would never have been able to achieve so much. Especially my parents (the best of the

world) who have always supported me and who keep trying to understand physics (poor people) and all what I do. And ALL my siblings, especially Eva and Alberto, for “los consejos de sabios” that helped me in the last time so much and I will always enjoy.

Selbständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschliesslich die angegebenen Quellen und Hilfen in Anspruchgenommen habe.

Berlin, den 28 Nov 2007 Ana Yagües